

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
13. März 2003 (13.03.2003)

PCT

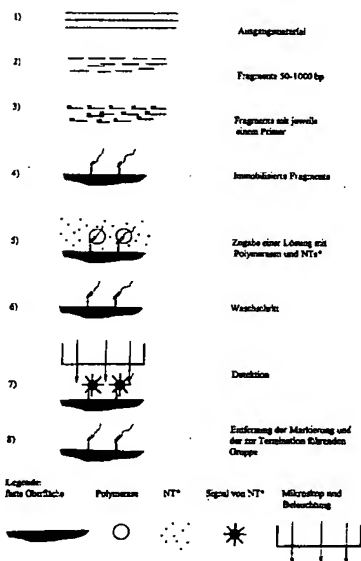
(10) Internationale Veröffentlichungsnummer  
WO 03/020968 A2

- (51) Internationale Patentklassifikation<sup>7</sup>: C12Q 1/68 (72) Erfinder; und  
(21) Internationales Aktenzeichen: PCT/EP02/09614 (75) Erfinder/Anmelder (nur für US): TCHERKASSOV,  
Dimitri [RU/DE]; Kahlhorststrasse 36, 23562 Lübeck  
(DE).  
(22) Internationales Anmeldedatum: 28. August 2002 (28.08.2002) (74) Anwälte: WEBER-QUITZAU, Martin usw.; Uexküll &  
Stolberg, Beselerstr. 4, 22607 Hamburg (DE).  
(25) Einreichungssprache: Deutsch  
(26) Veröffentlichungssprache: Deutsch (81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT,  
AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR,  
CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR ANALYZING NUCLEIC ACID SEQUENCES AND GENE EXPRESSION

(54) Bezeichnung: VERFAHREN ZUR ANALYSE VON NUKLEINSÄUREKETTENSEQUENZEN UND DER GENEXPRES-  
SION



(57) Abstract: The invention relates to a method for analyzing nucleic acid sequences and gene expression. The method is based on the detection of fluorescence signals of nucleotide molecules inserted into growing nucleic acid chains by means of a polymerase. The reaction occurs on a plane surface, wherein a plurality of individual nucleic acid molecules are immobilized on said surface. All said nucleic acid molecules are exposed to the same conditions so that a build-up reaction can simultaneously take place in all nucleic acid molecules.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Analyse von Nukleinsäurekettensequenzen und der Genexpression. Grundlage der Methode ist die Detektion von Fluoreszenzsignalen einzelner, mit Farbstoffen markierter Nukleotidmoleküle, die durch eine Polymerase in wachsende Nukleinsäureketten eingebaut werden. Die Reaktion verläuft auf einer planen Oberfläche. Auf dieser Oberfläche sind viele einzelne Nukleinsäure-Moleküle immobilisiert. Alle diese Nukleinsäure-Moleküle sind gleichen Bedingungen ausgesetzt, so dass an allen Nukleinsäure-Molekülen gleichzeitig eine Aufbaureaktion ablaufen kann.

1) STARTING MATERIAL  
2) FRAGMENTED 50-1000 BP  
3) FRAGMENTED WITH A PRIMER  
4) IMMOBILIZED FRAGMENTS  
5) ADDITION OF A SOLUTION WITH POLYMERASE AND NT\*  
6) WASHING STEP  
7) DETECTION  
8) REMOVAL OF THE MARKING AND OF THE GROUP LEADING TO TERMINATION

LEGEND  
- SOLID SURFACE  
- POLYMERASE  
- NT\*  
- SIGNAL FROM NT\*  
- MICROSCOPY AND ILLUMINATION

WO 03/020968 A2



(84) **Bestimmungsstaaten (regional):** ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Veröffentlicht:**

— *ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts*

*Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.*

## Verfahren zur Analyse von Nukleinsäurekettensequenzen und der Genexpression

Die Erfindung betrifft ein Verfahren zur Analyse von Nukleinsäureketten und der Genexpression. Die Grundlage der Methode ist die Detektion von Fluoreszenzsignalen einzelner mit Farbstoffen markierter Nukleotidmoleküle, die durch eine Polymerase in eine wachsende Nukleinsäurekette eingebaut werden. Die Reaktion verläuft auf einer planen Oberfläche. An diese Oberfläche sind viele einzelne Nukleinsäure-Moleküle gebunden. Alle diese Nukleinsäure-Moleküle sind gleichen Bedingungen ausgesetzt, so dass an allen Nukleinsäure-Molekülen gleichzeitig eine Aufbaureaktion ablaufen kann.

Das Verfahren umfaßt im wesentlichen folgende Schritte:

- 1) Bindung der Nukleinsäurekettenfragmenten (NSKFs) auf einer planen Oberfläche mit anschließender Hybridisierung von Primern, alternativ Bindung von Primern mit anschließender Hybridisierung von NSKFs, so dass NSKF-Primer-Komplexe gebildet werden.
- 2) Durchführen einer zyklischen Aufbaureaktion, wobei jeder Zyklus aus folgenden Schritten besteht:
  - a) Zugabe einer Lösung mit markierten Nukleotiden (NTs\*) und Polymerase zu den gebundenen NSKF-Primer-Komplexen,
  - b) Inkubation der gebundenen NSKF-Primer-Komplexe mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
  - c) Waschen,
  - d) Detektion der Signale von einzelnen Molekülen,
  - e) Entfernung der Markierung von den eingebauten Nukleotiden,
  - f) Waschen.Gegebenenfalls erfolgen mehrfache Wiederholungen des Zyklus.
- 3) Analyse der detektierten Signale der einzelnen Moleküle.
- 4) Rekonstruktion der Sequenzen aus den Einzeldaten.

## 1. Abkürzungen und Begriffserläuterungen

DNA - Desoxyribonukleinsäure verschiedenen Ursprungs und  
5 unterschiedlicher Länge (genomische DNA, cDNA, ssDNA, dsDNA)

RNA - Ribonukleinsäure (meist mRNA)

Polymerasen - Enzyme, die komplementäre Nukleotide in einen  
0 wachsenden DNA- oder RNA-Strang einbauen können ( z.B. DNA-  
Polymerasen, Reverse-Transkriptasen, RNA-Polymerasen)

dNTP - 2'-deoxy-Nucleosid-Triphosphate, Substrate für DNA-  
Polymerasen und Reverse-Transkriptasen  
5

NTP - Nukleosid-Triphosphate, Substrate für RNA-Polymerasen

NT - natürliches Nukleotid, meist dNTP, wenn nicht  
ausdrücklich anders gekennzeichnet.  
0

Abkürzung "NT" wird auch bei der Längenangabe einer  
Nukleinsäuresequenz verwendet, z.B. 1.000 NT. In diesem Fall  
steht "NT" für Nukleosid-Monophosphate.

5 Im Text wird bei Abkürzungen die Mehrzahl durch Verwendung des  
Suffixes "s" gebildet, "NT" steht zum Beispiel für  
"Nukleotid", "NTs" steht für mehrere Nukleotide.

NT' - modifiziertes Nukleotid, meist dNTP, wenn nicht  
0 ausdrücklich anders gekennzeichnet. NTs' bedeutet:  
modifizierte Nukleotide

NSK - Nukleinsäurekette. DNA oder RNA in ihrer ursprünglichen  
Länge  
5

NSKF - Nukleinsäurekettenfragment (DNA oder RNA), das einem  
Teil der Gesamtsequenz entspricht, NSKFs -



Nukleinsäurekettenfragmente. Die Summe der NSKFs bildet ein Äquivalent zur Gesamtsequenz. Die NSKFs können beispielsweise Fragmente von DNA- oder RNA-Gesamtsequenz sein, die nach einem Fragmentierungsschritt entstehen.

5

Gesamtsequenz - die in der Sequenzierungsreaktion eingesetzte Sequenz oder die eingesetzten Sequenzen, meistens in NSKFs überführt. Sie kann ursprünglich aus einer oder mehreren NSKs bestehen. Dabei kann die Gesamtsequenz Teile oder Äquivalente einer anderen Sequenz oder von Sequenz-Populationen darstellen (z.B. mRNA, cDNA, Plasmid-DNA mit Insert, BAC, YAC) und aus einer oder unterschiedlichen Spezies stammen.

0

Primerbindungstelle (PBS) - Teil der Sequenz in der NSK oder NSKF, an den der Primer bindet.

5

Referenzsequenz - eine bereits bekannte Sequenz, zu der die Abweichungen in der zu untersuchenden Sequenz bzw. in den zu untersuchenden Sequenzen (Gesamtsequenz) ermittelt werden.

Als Referenzsequenzen können in Datenbanken zugängliche Sequenzen verwendet werden, wie z.B. aus der NCBI-Datenbank.

0

$T_m$  - Schmelztemperatur

Plane Oberfläche - Oberfläche, die vorzugsweise folgende Merkmale aufweist: 1) Sie erlaubt, mehrere einzelne Moleküle, vorzugsweise mehr als 100, noch besser mehr als 1000, mit dem jeweiligen gegebenen Objektiv-Oberfläche-Abstand bei einer Objektivposition gleichzeitig zu detektieren. 2) Die immobilisierten einzelnen Moleküle befinden sich in derselben Fokusebene, die reproduzierbar eingestellt werden kann.

5

0

Weitfeld-Optik-Detektionssystem - Detektionssystem, das gleichzeitig Fluoreszenzsignale von einzelnen, auf einer Fläche verteilten Molekülen detektieren kann, wobei die Fläche ca.  $100 \mu m^2$  und größer ist. Ein Beispiel für Weitfeld-Detektionsoptik stellt Fluoreszenzmikroskop Axiovert 200 oder

5

Axioplan 2e (Zeiss) mit einem Planneofluar-Objektiv 100x NA 1.4 Ölimmersion (Zeiss), oder einem Planapochromat-Objektiv 100x NA 1.4 Ölimmersion (Zeiss); die Anregung der Fluoreszenz kann dabei mit einer Lampe, z.B. Quecksilberdampf Lampe, oder einem Laser oder Dioden erfolgen. Sowohl Epifluoreszenzmodus als auch im Totalreflexions-Fluoreszenzmikroskopie-Modus (total internal reflection fluorescence microscopy, TIRF-Microscopy) oder Laser-Scanning-Mikroskopie-Modus können verwendet werden. In dieser Anmeldung wird Gebrauch von dieser Weitfeld-Detektionsoptik gemacht

Definition der Termination: Als Termination wird in dieser Anmeldung der reversible Stop des Einbaus der modifizierten ungespalteten NTs\* bezeichnet.

Dieser Begriff ist von dem üblichen Gebrauch des Wortes "Termination" durch Dideoxy-NTP bei einer konventionellen Sequenzierung zu trennen.

Die Termination kommt nach dem Einbau eines modifizierten NT\* zustande. Ein zur Termination führender Substituent- bzw. eine Modifikation der 3'-OH-Position an der Desoxyribose eines Nukleotides, die zur Termination führt. Der Substituent kann unter milden Bedingungen abgespalten werden, so daß 3'-OH Funktion wieder für den Einbau eines NT\* zur Verfügung steht. An diesen Substituenten ist ein Fluoreszenzfarbstoff gekoppelt.

Genprodukte - Bei den Genprodukten handelt es sich um die primären Genprodukte der Gene. Im wesentlichen handelt es sich dabei um RNA-Transkripte der genannten Gene, welche auch als Target-Sequenzen (oder Target-Nukleinsäuresequenzen) bezeichnet werden. Diese Target-Sequenzen schließen neben mRNA auch davon abgeleitete einzelsträngige und doppelsträngige cDNA, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA ein.

Einzelnukleotidpolymorphismen (single nucleotide

polymorphisms, SNPs) - Veränderungen in den Sequenzen, die als Substitution (Transition oder Transversion) oder als Deletion oder Insertion einzelner NT auftreten können.

## 2. Stand der Technik

Die Nukleinsäurenketten-Sequenzanalyse und Genexpressionsanalyse sind in vielen Bereichen der Wissenschaft, Medizin und Industrie zu einem wichtigen Werkzeug geworden. Zur Analyse wurden mehrere Verfahren entwickelt.

Die bekanntesten Verfahren sind die Ketten-Terminations-Sequenzierung nach Sanger (F. Sanger et al. PNAS 1977 v.74 s. 5463), die auf dem Einbau von Kettenterminatoren basiert, und die Maxam-Gilbert-Methode, die auf Basen-spezifischer Modifikation und Spaltung von Nukleinsäureketten beruht (A.M. Maxam and W. Gilbert PNAS 1977, v.74 S.560). Beide Methoden liefern eine Anzahl von Nukleinsäurekettenfragmenten verschiedener Längen. Diese Fragmente werden der Länge nach in einem Gel aufgetrennt. Dabei müssen alle Nachteile der Elektrophorese (wie z.B. lange Laufzeit, relativ kurze Strecken von Sequenzen, die in einem Ansatz bestimmt werden können, begrenzte Anzahl der parallelen Ansätze sowie relativ große Mengen an DNA) in Kauf genommen werden. Diese Methoden sind sehr arbeitsintensiv und langsam.

Ein weiteres Verfahren zur Sequenzierung basiert auf der Hybridisierung von Nukleinsäureketten mit kurzen Oligonukleotiden. Dabei wird mit mathematischen Methoden berechnet, wie viele Oligonukleotide einer bestimmten Länge vorhanden sein müssen, um eine komplette Sequenz zu ermitteln (Z.T. Strezoska et al. PNAS 1991 v.88 S.10089, R.S.Drmanac et al. Science 1993 v.260 S.1649). Auch dieses Verfahren ist mit Problemen behaftet: Es kann nur eine Sequenz in einem Ansatz bestimmt werden, sekundäre Strukturen stören die

Hybridisierung und Sequenzwiederholungen verhindern die korrekte Analyse.

Eine andere Möglichkeit zur Sequenzierung haben Arbeitsgruppen beispielsweise von (Dower US Patent 5.547.839, Canard et al. US Patent 5.798.210, Rasolonjatovo Nucleosides & Nucleotides 1999, v.18 S.1021, Metzker et al. NAR 1994, v.22, S.4259, Welch et al. Nucleosides & Nucleotides 1999, v.18, S.197) entwickelt. Diese Methode wird abgekürzt als BASS (Base Addition Sequencing Scheme) oder SBS (Sequencing by Synthesis) bezeichnet. Dabei wird eine große Anzahl gleicher einzelsträngiger DNA-Stücke an einem definierten Ort auf einer Oberfläche fixiert und das Signal von der Gesamtheit dieser vielen identischen DNA-Stücke analysiert. Zu dieser fixierten DNA wird eine Lösung mit Polymerase und Nukleotiden zugegeben, so dass ein komplementärer Strang synthetisiert werden kann. Dabei soll die Polymerase schrittweise arbeiten: in jedem Schritt wird nur ein einziges Nukleotid eingebaut. Dieses wird detektiert, worauf die Polymerase in einem nächsten Zyklus das nächste Nukleotid einbaut.

Trotz des Gelingens einiger einzelner Schritte der Methode wurde sie nicht zu einem funktionsfähigen Verfahren entwickelt. Dies kann beispielsweise auf folgenden Tatsachen beruhen: Beim Aufbau der komplementären Stränge tritt sehr schnell eine Desynchronisation der Synthese auf, so dass bei jedem Schritt die Fehler akkumulieren. Deshalb können nur sehr kurze Fragmente sequenziert werden. Es ist zu betonen, dass alle beschriebenen BASS-Methoden nicht auf der Detektion von einzelnen Molekülen beruhen. Das Signal wird stattdessen von einer großen Anzahl identischer an einem definierten Ort immobilisierter Moleküle registriert. Die in diesen Methoden übliche Verwendung der Begriffe "einzelne Moleküle" und "Moleküle" zielt dabei nicht auf individuelle, voneinander getrennte Moleküle, sondern auf eine Population, die aus vielen identischen Molekülen besteht. Identisch heißt in diesem Fall, dass die Moleküle die gleiche Sequenz haben.

Die Analyse des Genexpressions-Spektrums ist zu einem wichtigen Werkzeug in der Wissenschaft geworden. Der Vergleich der Genexpressions-Spektren zwischen verschiedenen Zelllinien, Geweben oder Entwicklungsstadien erlaubt Rückschlüsse auf die darin ablaufenden spezifischen biologischen Prozesse. So kann man z.B. erwarten, dass der Vergleich zwischen Tumorzellen und gesunden Zellen gleicher Herkunft Auskunft über die am Tumorgeschehen beteiligten Gene gibt. Dabei ist wichtig, dass die Aktivität möglichst vieler oder aller Gene gleichzeitig analysiert wird.

Die Analyse der Genexpression ist eine komplexe Aufgabe: Die Anzahl der in einem Zelltyp aktiven Gene kann mehrere Tausend betragen. Die Analyse sollte aber möglichst alle im Genom der betreffenden Art enthaltenen Gene (etwa 32000 beim Menschen) berücksichtigen. Hinzu kommt, dass die im jeweiligen Zelltyp aktiven Gene erstens meist noch nicht komplett bekannt sind und zweitens unterschiedlich stark exprimiert werden.

Es wurden bereits viele Methoden zur Genexpressionsanalyse entwickelt, so z.B. Differential Display (Nature 1984, v.308 S.149, Science 1992 v.257 S.967), Expressed Sequence Tags (EST) (Science 1991 v.252, S.1656, Nature Genetics 1992, v.2 S.173), Northern blotting oder RT-PCR (PNAS 1977, v.74, S.5350, Cell 1983 v.34 S.865, "The PCR Technique, RT-PCR" 1998, Ed. Paul Swebert, Eaton Publishing). Alle diese Methoden können nur eine sehr begrenzte Anzahl an Genen pro Reaktion analysieren und sind zum Teil sehr arbeitsintensiv.

Die am weitesten verbreitete Methode zur parallelen Analyse der Genexpressionsmuster ist die Hybridisierung eines zu analysierenden Gemischs von cDNA-Molekülen mit an eine Oberfläche gebundenen Oligonukleotiden, die in einer bestimmten Anordnung, meist als "Microarray" fixiert sind ("Microarray Biochip Technology" 2000, Ed. M.Schena, Eaton Publishing, Zhao et al. Gene 1995, v. 156, S.207, Schena et al. Science 1995 v.270, S.467, Lockhart et al. US Patent

6.040.138, Wang US Patent 6.004.755, Arlinghaus et al. US Patent 5.821.060, Southern US Patent 5.700.637, Fodor et al. US Patent 5.871.928).

Zu den großen Nachteilen der Hybridisierungsmethode zählen: Die Fertigung der an die Oberfläche gebundenen Oligonukleotide ist teuer. Die Analyse beschränkt sich auf Gene, deren Sequenzen bereits bekannt sind. Mehrere Mismatch-Kontrollen vergrößern die Anzahl der Oligonukleotide, die immobilisiert werden müssen.

Die Aufgabe der vorliegenden Erfindung besteht daher darin, ein Verfahren zur Sequenzanalyse von Nukleinsäureketten und der Analyse der Genexpression bereitzustellen, das die Nachteile der oben erwähnten Methoden nicht aufweist und vor allem eine billigere, schnellere und effizientere Analyse von Nukleinsäuresequenzen ermöglicht. Insbesondere soll das Verfahren in der Lage sein, viele Sequenzen parallel zu bestimmen. Es kann dann beispielsweise für die Analyse sehr langer Nukleinsäureketten (mehrere Mb) oder für die Variantenanalyse an vielen kurzen Ketten (Mutationsanalyse, SNP-Analyse) in einem Ansatz verwendet werden.

### 3. Kurze Beschreibung

Gegenstand der vorliegenden Erfindung ist ein Verfahren zur parallelen Sequenzanalyse von Nukleinsäuresequenzen (Nukleinsäureketten, NSKs) gelöst, bei dem man

Fragmente (NSKFs) einzelsträngiger NSKs mit einer Länge von etwa 50 bis 1000 Nukleotiden erzeugt, die überlappende Teilsequenzen der Gesamtsequenzen darstellen, man

die NSKFs unter Verwendung eines einheitlichen oder mehrerer unterschiedlicher Primer in Form von NSKF-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der NSKFs unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

- a) zu den an die Oberfläche gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs\*) enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs\* jeweils an den NTs\* befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs\* durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs\* strukturell so modifiziert sind, daß die Polymerase nach Einbau eines solchen NT\* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT\* in denselben Strang einzubauen, wobei der zur Termination führende Substituent mit dem Fluoreszenzfarbstoff abspaltbar ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT\* verlängert werden, man
- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs\* geeignet sind, man
- d) die einzelnen, in komplementäre Stränge eingebauten NTs\* durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position

der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man

- e) zur Erzeugung unmarkierter (NTs oder) NSKFs die zur Termination führenden Substituenten und die Fluoreszenzfarbstoffe von den am komplementären Strang angefügten NTs\* abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

Aus den ermittelten Teilsequenzen kann man beispielsweise die Gesamtsequenz der NSKs bestimmen. Unter einer parallelen Sequenzanalyse wird in diesem Zusammenhang die gleichzeitige Sequenzanalyse vieler NSKFs verstanden (beispielsweise 1.000.000 bis 10.000.000), wobei diese NSKFs von einer einheitlichen NSK-Population oder von mehreren unterschiedlichen NSK-Populationen abgeleitet sind.

Die erhaltene Population von überlappenden Teilsequenzen läßt sich beispielsweise bei *de novo* Sequenzierung mit kommerziell erhältlichen Programmen zur Gesamtsequenz der NSK zusammenfügen (Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257).

Bei der Analyse von Varianten einer bekannten Referenzsequenz lassen sich Mutationen oder Einzelnukleotidpolymorphismen



durch einen Vergleich der erhaltenen überlappenden Teilsequenzen mit der Referenzsequenz feststellen.

Gemäß einer besonderen Ausführungsform der Erfindung kann das Verfahren durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man

- a) in jedem Zyklus nur jeweils ein markiertes NT<sup>\*</sup>,
- b) in jedem Zyklus jeweils zwei unterschiedlich markierte NTs<sup>\*</sup> oder
- c) in jedem Zyklus jeweils vier unterschiedlich markierte NTs<sup>\*</sup>

einsetzt.

Wenn die NSKs Varianten einer bekannten Referenzsequenz sind kann das Verfahren auch durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei unterschiedlich markierte NTs<sup>\*</sup> und zwei unmarkierte NTs einsetzt und man die Gesamtsequenzen durch Vergleich mit der Referenzsequenz ermittelt.

Gegenstand der vorliegenden Erfindung ist ferner ein Verfahren zur hoch parallelen Analyse der Genexpression, bei dem man

einzelsträngige Genprodukte bereitstellt, man

die Genprodukte unter Verwendung eines einheitlichen oder mehrerer unterschiedlichen Primer in Form von Genprodukt-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der Genprodukte unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

- a) zu den auf der Oberfläche gebundenen Genprodukt-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs' strukturell so modifiziert sind, daß die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der zur Termination führende Substituent mit dem Fluoreszenzfarbstoff abspaltbar ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man
- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs' geeignet sind, man
- d) die einzelnen, in komplementäre Stränge eingebauten NTs' durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- e) zur Erzeugung unmarkierter (NTs oder) NSKFs die zur Termination führenden Substituenten mit den

Fluoreszenzfarbstoffen von den am komplementären Strang angefügten NTs abspaltet, man

- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner Genprodukt-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser Genprodukte durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt und man aus den ermittelten Teilssequenzen die Identität der Genprodukte bestimmt.

Bei den Genprodukten handelt es sich um die primären Genprodukte der Gene, deren Expression analysiert werden soll. Im wesentlichen handelt es sich dabei um RNA-Transkripte der genannten Gene, welche auch als Target-Sequenzen (oder Target-Nukleinsäuresequenzen) bezeichnet werden. Diese Target-Sequenzen schließen neben mRNA auch davon abgeleitete einzelsträngige und doppelsträngige cDNA, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA ein.

Die Genprodukte oder Target-Sequenzen können entweder als mRNAs direkt aus einer biologischen Probe (z.B. Zellextrakt, Gewebeextrakt oder Extrakt von ganzen Organismen) isoliert oder als cDNAs durch reverse Transkription der mRNAs erhalten werden.

Unter einer hoch parallelen Analyse wird in diesem Zusammenhang die gleichzeitige Sequenzanalyse vieler Genprodukt-Moleküle verstanden (beispielsweise 1.000.000 bis 10.000.000), wobei diese Genprodukt-Moleküle eine komplexe

heterogene Population darstellen, die z.B. einem kompletten Expressionsprofil bzw. einem Expressionsspektrum eines Gewebes entspricht.

Gemäß einer besonderen Ausführungsform der Erfindung kann das Verfahren durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man

- a) in jedem Zyklus nur jeweils ein markiertes NT\*,
- b) in jedem Zyklus jeweils zwei unterschiedlich markierte NTs\* oder
- c) in jedem Zyklus jeweils vier unterschiedlich markierte NTs\*

einsetzt.

Das Verfahren kann auch durchgeführt werden, indem man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei unterschiedlich markierte NTs\* und zwei unmarkierte NTs einsetzt und man die Identität der Genprodukte durch Vergleich mit den Referenzsequenzen ermittelt.

Gegenstand der Erfindung ist ferner ein Kit zur Durchführung des Verfahrens das eine Reaktionsoberfläche, zur Durchführung des Verfahrens erforderliche Reaktionslösungen, eine oder mehrere Polymerasen, und Nukleotide (NTs) enthält, von denen ein bis vier mit Fluoreszenzfarbstoffen markiert sind, wobei die NTs\* an der 3'-Position strukturell so modifiziert sind, daß die Polymerase nach Einbau eines solchen NT\* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT\* in denselben Strang einzubauen, wobei der zur Termination führende Substituent mit dem Fluoreszenzfarbstoff abspaltbar ist. Gemäß einer besonderen Ausführungsform der Erfindung enthält das Kit ferner zur Erzeugung von

Einzelsträngen aus Doppelsträngen erforderliche Reagenzien, einzelsträngige Nukleinsäuremoleküle, die als PBS in die NSKFs eingeführt werden, Oligonukleotid-Primer, zur Abspaltung der Fluoreszenzfarbstoffe und der zur Termination führenden Substituenten erforderliche Reagenzien und/oder Waschlösungen.

Die erfindungsgemäße Methode dient zur Ermittlung der Nukleinsäuresequenzen und kann in verschiedenen Bereichen der Genetik eingesetzt werden. Dazu zählen insbesondere die Bestimmung unbekannter, langer Sequenzen, Analysen von Sequenz-Polymorphismen und Punktmutationen sowie die parallele Analyse einer großen Zahl an Gensequenzen, sowie die Analyse der Genexpression.

Bei der Analyse langer Nukleinsäureketten (z.B. 100 Kb und länger) hängt die Vorbereitung des zu analysierenden Materials (einzeln- und doppelsträngige Nukleinsäuresequenzen) von der Aufgabestellung ab und hat das Ziel, aus einer langen Nukleinsäurekette eine Population an relativ kleinen, einzelsträngigen Nukleinsäurekettenfragmenten (NSKFs) zu bilden, diese Fragmente mit einem für den Start der Sequenzierungsreaktion geeigneten Primer zu versehen (NSKF-Primer-Komplexe) und auf einer planen Oberfläche zu fixieren.

Dabei werden einzelne NSKFs auf einer planen Oberfläche in einer solchen Weise fixiert, dass eine enzymatische Reaktion an diesen Molekülen ablaufen kann. Prinzipiell sind verschiedene Arten der Immobilisation möglich, die von der Zielsetzung, der Art der NSK und der für die Reaktion eingesetzten Polymerase abhängen. Die NSKFs werden bei der Immobilisierung bzw. Bindung zufällig auf der Oberfläche verteilt, d.h. es muß also nicht auf eine exakte Positionierung der einzelnen Ketten geachtet werden. NSKF-Primer-Komplexe können über die NSKFs oder Primer an die Oberfläche gebunden werden. Die NSKF-Primer-Komplexe müssen dabei in einer solchen Dichte auf der Oberfläche fixiert

werden, dass eine eindeutige Zuordnung der später detektierten Signale von den eingebauten NT\*s zu einzelnen NSKFs gewährleistet ist.

Nach der Vorbereitung der NSKFs startet man mit allen auf der Oberfläche immobilisierten NSKF-Primer-Komplex-Molekülen die Sequenzierungsreaktion. Als Grundlage der Sequenzierung dient die Synthese des komplementären Stranges zu jedem einzelnen gebundenen NSKF. Dabei werden in den neu synthetisierten Strang markierte NTs\* eingebaut. Die Polymerase baut in einem Zyklus nur ein einziges markiertes NT\* in die wachsende Kette ein.

Die Sequenzierungsreaktion verläuft in mehreren Zyklen. Ein Zyklus umfasst folgende Schritte:

- a) Zugabe einer Lösung mit markierten Nukleotiden (NTs\*) und Polymerase zu den gebundenen NSKF-Primer-Komplexen,
- b) Inkubation der gebundenen NSKF-Primer-Komplexe mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- c) Waschen,
- d) Detektion der Signale von einzelnen Molekülen,
- e) Entfernung der Markierung von den eingebauten Nukleotiden,
- f) Waschen.

Gegebenenfalls erfolgt eine mehrfache Wiederholung des Zyklus (a-f).

Die Reaktionsbedingungen des Schrittes (b) in einem Zyklus werden so gewählt, dass die Polymerasen an mehr als 50% der an der Sequenzierungsreaktion beteiligten NSKFs (extensionsfähige NSKF-Primer-Komplexe) in einem Zyklus ein markiertes NT\* einbauen können, vorzugsweise an mehr als 90%.

Die Anzahl der durchzuführenden Zyklen hängt dabei von der jeweiligen Aufgabenstellung ab, ist theoretisch nicht beschränkt

und liegt vorzugsweise zwischen 20 und 5000.

Danach wird für jedes fixierte NSKF seine spezifische Sequenz aus der Reihenfolge der eingebauten NTs ermittelt.

Aus den überlappenden NSKF-Sequenzen kann in einer Ausführungsform die ursprüngliche NSK-Sequenz rekonstruiert werden ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Dabei sucht man in der gesamten Population von NSKF-Sequenzen nach Übereinstimmungen/Überlappungen in den Sequenzen von NSKFs. Durch diese Übereinstimmungen/Überlappungen kann man die NSKF in eine Reihe bringen, z.B.:

ACTGTGCGTCCGTATGATGGTCATTCCATG

CATTCCATGGTACGTTAGCTCCTAG

TCCTAGTAAAATCGTACC.

In der Praxis hat sich bei einer Sequenzierung von unbekannten Sequenzen bewährt, eine Länge der sequenzierten Stücke von mehr als 300 bp zu erreichen. Das erlaubt die Sequenzierung von Genomen aus Eukaryonten im Schrotschuss-Verfahren.

Dabei können die Fehler der Methode mit verschiedenen Mitteln erfasst und korrigiert werden. Sämtliche Schritte des Verfahrens können weitgehend automatisiert werden.

Durch die Arbeit mit einzelnen Molekülen ergeben sich große Vorteile gegenüber der früher beschriebenen BASS-Methode:

1. Da die Moleküle einzeln detektiert werden, besteht keine Gefahr, dass das Signal durch die Desynchronisation in der Population fehlerhaft wird. Für jedes fixierte NSKF wird

eine eigene Sequenz erstellt. Daher spielt es keine Rolle, ob an einem benachbarten Molekül die Synthese bereits weiter fortgeschritten oder zurückgeblieben ist. Dadurch wird eine hoch parallele Sequenzierung langer NSKF erst möglich.

2. Es ist nicht notwendig, Moleküle in einer definierten Anordnung auf der Oberfläche zu fixieren, da das Signal von einzelnen Molekülen ausgeht und nicht von einer räumlich definierten Population (was bei BASS-Methoden notwendig ist).
3. Es ist nicht zwingend notwendig multiple Kopien von den zu analysierenden Nukleinsäureketten herzustellen, so dass PCR- und bzw. oder Klonierungsschritte entfallen können. Dies führt zu einer enormen Beschleunigung der Analyse im Vergleich zu bestehenden Verfahren.

Das erfindungsgemäße Verfahren zur Analyse der Genexpression erhält durch eine gleichzeitige Sequenzierung einzelner Genproduktmoleküle mehrere Vorteile gegenüber bekannten Methoden der Analyse der Genexpression:

- 1) Die Genprodukte können in einer beliebigen Anordnung auf der Oberfläche binden. Eine vorherige aufwendige Synthese von verschiedenen Oligonukleotiden an bestimmten Positionen (wie beispielsweise bei der Hybridisierungsmethode) ist somit nicht notwendig.
- 2) Das Material kann auf einer standardisierten Oberfläche analysiert werden.
- 3) Auch die Expression noch unbekannter Gene kann ermittelt werden, weil alle im Ansatz enthaltenen Genprodukte analysiert werden.
- 4) Die große Anzahl der analysierten Moleküle erlaubt auch die Detektion schwach exprimierter Gene.
- 5) Kleinste Mengen an Ausgangsmaterial können eingesetzt werden: mRNA aus einer einzelnen Zelle kann für die Analyse ausreichend sein.



6) Sämtliche Schritte des Verfahrens können weitgehend automatisiert werden.

Die Methode basiert auf mehreren Prinzipien:

1. Kurze Nukleotidsequenzen (10-50 NTs) enthalten genügend Informationen zur Identifizierung des korrespondierenden Gens, wenn die Gensequenz selbst bereits in einer Datenbank enthalten ist.

Eine Sequenz aus beispielsweise 10 NTs kann mehr als  $10^6$  verschiedene Kombinationen bilden. Das ist z.B. für die meisten Gene im menschlichen Genom, das nach heutiger Schätzung 32000 Gene enthält, ausreichend. Für Organismen mit weniger Genen kann die Sequenz noch kürzer sein.

2. Der Methode liegt ein neues Verfahren zur die Sequenzierung einzelner Nukleinsäurekettenmoleküle zugrunde.

3. Es können Nukleinsäureketten-Gemische untersucht werden.

4. Die Sequenzierungsreaktion läuft an vielen Molekülen gleichzeitig ab, wobei die Sequenz jeder einzelnen gebundenen Nukleinsäurekette analysiert wird.

Es ist bekannt, dass zur Untersuchung der Genexpression mRNAs oder von der mRNA abgeleitete Nukleinsäureketten (z.B. einzelsträngige cDNAs, doppelsträngige cDNAs, von cDNA abgeleitete RNA oder von cDNA amplifizierte DNA) eingesetzt werden kann. Unabhängig von der genauen Zusammensetzung werden sie im folgenden als Genprodukte bezeichnet. Auch Teilsequenzen dieser Genprodukte werden im folgenden als Genprodukte bezeichnet.

Diese Genprodukte stellen ein Gemisch aus verschiedenen Nukleinsäureketten dar.

Als Grundlage der Analyse dient die Synthese eines zum Genprodukt komplementären Stranges.

Das Ziel der Vorbereitung ist, auf einer planen Oberfläche in zufälliger Weise gebundene Genprodukt-Primer-Komplexe bereitzustellen, an denen der Einbau von NT\*s durch die Polymerase stattfinden kann (extensionsfähige Genprodukt-Primer-Komplexe).

Mit diesen gebundenen Genprodukt-Primer-Komplexen wird die Sequenzierungsreaktion durchgeführt.

Sie verläuft in mehreren Zyklen. Pro Zyklus wird jeweils nur ein einziges markiertes NT\* in den wachsenden Strang eingebaut. Ein Zyklus umfasst folgende Schritte:

- a) Zugabe einer Lösung mit markierten Nukleotiden (NTs\*) und Polymerase zu gebundenen Genprodukt-Primer-Komplexen,
- b) Inkubation der gebundenen Genprodukt-Primer-Komplexe mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- c) Waschen,
- d) Detektion der Signale von einzelnen modifizierten, in die neu synthetisierten Stränge eingebauten NTs\*-Molekülen,
- e) Entfernung der Markierung von den eingebauten Nukleotiden,
- f) Waschen.

Dieser Zyklus kann mehrmals wiederholt werden, so dass von jedem an der Sequenzierungsreaktion teilnehmenden Genprodukt-Primer-Komplex vorzugsweise 10 bis 50 NTs ermittelt werden. Danach erfolgt die Rekonstruktion der Nukleinsäuresequenzen aus den detektierten Signalen. Die ermittelten Sequenzen der gebundenen Genprodukte werden zur Bestimmung der Abundanzen untereinander verglichen und durch Vergleich mit Gensequenzen in Datenbanken bestimmten Genen zugeordnet.

#### 4. Detaillierte Beschreibung

Allgemeine Prinzipien der Reaktion, Materialauswahl und Materialvorbereitung (Erzeugung kurzer NSKFs, Einführung einer PBS, Einzelstrangvorbereitung, Primerauswahl, Fixierung von NSKFs), sowie die Detektionsapparatur und Detektion

werden am Beispiel des Verfahrens zur Sequenzierung langer NSKs dargestellt. Das Verfahren zur Analyse der Genexpression wird anschließend im Beispiel 3 beschrieben.

#### 4.1 Allgemeine Prinzipien der Reaktion

Im folgenden sollen anhand der Sequenzierung eines mehrere Mb langen DNA-Stückes beispielhaft die allgemeinen Prinzipien der Reaktion dargestellt werden (Fig. 1). Der Sequenzierung und der Rekonstruktion von Nukleinsäuresequenzen liegt das Shotgun-Prinzip zugrunde ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Die Sequenz eines langen DNA-Stücks wird dabei durch die Sequenzierung kleiner DNA-Fragmente und nachfolgender Rekonstruktion ermittelt. Das zu analysierende Material (1) wird für die Sequenzierungsreaktion vorbereitet, indem es in Fragmente von vorzugsweise 50 bis 1000 bp Länge zerlegt wird (2). Jedes Fragment wird anschließend mit einer Primerbindungsstelle und einem Primer versehen (3). Dieses Gemisch aus verschiedenen DNA-Fragmenten wird nun auf einer planen Oberfläche fixiert (4). Die nicht gebundenen DNA-Fragmente werden durch einen Waschschrift entfernt. Danach wird die Sequenzierungsreaktion an der gesamten Reaktionsoberfläche durchgeführt. Diese Reaktion verläuft zyklisch. Im 1. Schritt des Zyklus wird ein mit einem Fluoreszenzfarbstoff markiertes NT<sup>\*</sup> in den wachsenden Strang eingebaut: Dabei wird die Reaktion so gesteuert, dass in jedem Zyklus jeweils nur ein markiertes NT<sup>\*</sup> von einer Polymerase in den wachsenden Strang eingebaut werden kann. Das wird durch die Verwendung von NTs<sup>\*</sup> erreicht, die an der 3'-Position der Desoxyribose ein reversibel gekoppeltes, zur Termination führenden Substituenten tragen. Der Einbau eines weiteren markierten NT<sup>\*</sup> wird dadurch unmöglich gemacht. Die Polymerase und die markierten NTs<sup>\*</sup> werden gleichzeitig in die Reaktion eigesetzt (5). Danach wird das Reaktionsgemisch entfernt und

die Oberfläche in geeigneter Art und Weise gewaschen (6). Nun folgt ein Detektionsschritt (7): Die Oberfläche wird mit einer für die Einzelmoleküldetektion geeigneten Vorrichtung (bestehend aus Lichtquelle, Mikroskop, Kamera, Scantisch, Computer mit Steuerungs- und Bilderkennungs- bzw. Bildverarbeitungssoftware) abgescannt und die Signale der einzelnen, eingebauten markierten NTs\* identifiziert. Nach dem Detektionsschritt wird die Markierung und der zur Termination führende Substituent von allen eingebauten NTs\* entfernt (8). Nach einem sich anschließenden Waschschrift kann ein neuer Zyklus beginnen. Zur Rekonstruktion einer größeren ursprünglichen DNA-Sequenz (z.B. mehrere Mb langes DNA-Stück) sollen die DNA-Fragmente einige Hundert NT lang sein, falls man die Rekonstruktion nach dem Shotgun-Prinzip durchführt ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Da pro Zyklus nur jeweils ein markiertes NT\* eingebaut wird, sind mindestens 300 Zyklen zur Sequenzierung notwendig.

#### 4.2 Auswahl des Materials

Mit Hilfe der erfindungsgemäßen Methode ist es möglich, sowohl vorselektionierte DNA-Sequenzen (z.B. in YAC-, PAC-, oder BAC-Vektoren (R. Anand et al. NAR 1989 v.17 S.3425, H. Shizuya et al. PNAS 1992 v.89 S.8794, "Construction of bacterial artificial chromosome libraries using the modified PAC system" in "Current Protocols in Human genetics" 1996 John Wiley & Sons Inc.) klonierte Abschnitte eines Genoms) als auch nicht vorselektionierte DNA (z.B. genomische DNA, cDNA-Gemische) zu analysieren. Durch eine Vorselektion ist es möglich, im Vorfeld relevante Informationen, wie z.B. Sequenz-Abschnitte aus einem Genom oder Populationen an Genprodukten, aus der große Menge genetischer Informationen herauszufiltern und damit die Menge der zu analysierenden Sequenzen einzuschränken. Besonders hervorzuheben sind die

Ausführungsformen, bei denen die erfindungsgemäßen Verfahren ohne Vorselektionierung und ohne eine Vervielfältigung des Materials eingesetzt werden. Verzicht auf PCR und Klonierung bringt eine entscheidende Beschleunigung in der hoch parallelen Analyse von Nukleinsäuresequenzen, was mit anderen Verfahren bis jetzt nicht möglich war.

#### 4.3 Vorbereitung des Materials

Ziel der Materialvorbereitung ist es, gebundene einzelsträngige NSKFs mit einer Länge von vorzugsweise 50-1000 NTs, einer einzelnen Primerbindungsstelle und einem hybridisierten Primer (gebundene NSKF-Primer-Komplexe) zu erhalten. Diese NSKF-Primer-Komplexe haben beispielsweise die in Fig. 2 dargestellte Struktur. Im einzelnen können sehr variable Konstruktionen aus dieser allgemeinen Struktur abgeleitet werden. Zur Verbesserung der Anschaulichkeit folgen nun einige Beispiele, wobei die angeführten Methoden einzeln oder in Kombination eingesetzt werden können.

##### 4.3.1 Erzeugung kurzer Nukleinsäurekettenfragmente (50-1000 NTs) (Fragmentierungsschritt)

Wichtig ist, dass die Fragmentierung der NSKs so erfolgt, dass Fragmente erhalten werden, die überlappende Teilsequenzen der Gesamtsequenzen darstellen. Dies wird durch Verfahren erreicht, bei denen unterschiedlich lange Fragmente als Spaltprodukte in zufallsmäßiger Verteilung entstehen.

Erfindungsgemäß kann die Erzeugung der Nukleinsäurekettenfragmente (NSKFs) durch mehrere Methoden erfolgen, z.B. durch die Fragmentierung des Ausgangsmaterials mit Ultraschall oder durch Endonukleasen ("Molecular cloning". 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press), wie z.B. durch unspezifische Endonukleasegemische. Erfindungsgemäß wird die Ultraschall-Fragmentierung bevorzugt. Man kann die Bedingungen so einstellen, dass Fragmente mit einer durchschnittlichen

Länge von 100 bp bis 1 kb entstehen. Diese Fragmente werden anschließend an ihren Enden durch das Klenow-Fragment (E.coli-Polymerase I) oder durch die T4-DNA-Polymerase aufgefüllt ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press).

Ausserdem können aus einer langen NSK unter Verwendung randomisierter Primer komplementäre kurze NSKFs synthetisiert werden. Besonders bevorzugt wird diese Methode bei der Analyse der Gen-Sequenzen. Dabei werden an der mRNA einzelsträngige DNA-Fragmente mit randomisierten Primern und einer reversen Transkriptase gebildet (Zhang-J et al. Biochem.J. 1999 v.337 S.231, Ledbetter et al. J.Biol.Chem. 1994 v.269 S.31544, Kolls et al. Anal.Biochem. 1993 v.208 S.264, Decraene et al. Biotechniques 1999 v.27 S.962).

#### 4.3.2 Einführung einer Primerbindungsstelle in das NSKF.

Die Primerbindungsstelle (PBS) ist ein Sequenzabschnitt, der eine selektive Bindung des Primers an das NSKF ermöglichen soll.

In einer Ausführungsform können die Primerbindungsstellen unterschiedlich sein, so dass mehrere unterschiedliche Primer verwendet werden müssen. In diesem Fall können bestimmte Sequenzabschnitte der Gesamtsequenz als natürliche PBSs für spezifische Primer dienen. Diese Ausführungsform ist besonders für die Untersuchung bereits bekannter SNP-Stellen geeignet, s. Beispiel 4 "SNP-Analyse mit sequenzspezifischen Primern".

In einer anderen Ausführungsform ist es aus Gründen der Vereinfachung der Analyse günstig, wenn eine einheitliche Primerbindungsstelle in allen NSKFs vorhanden ist. Gemäß einer bevorzugten Ausführungsform der Erfindung werden die Primerbindungsstellen daher in die NSKFs extra eingeführt. Auf diese Weise können Primer mit einheitlicher Struktur für die Reaktion eingesetzt werden.

Im folgenden wird diese Ausführungsform detailliert beschrieben.

Die Zusammensetzung der Primerbindungsstelle ist nicht eingeschränkt. Ihre Länge beträgt vorzugsweise zwischen 20 und 50 NTs. Die Primerbindungsstelle kann eine funktionelle Gruppe zur Immobilisation des NSKF tragen. Diese funktionelle Gruppe kann z.B. eine Biotingruppe sein.

Als Beispiel für die Einführung einer einheitlichen Primerbindungsstelle werden im folgenden die Ligation und das Nukleotid-Tailing an DNA-Fragmente beschrieben.

a) Ligation:

Dabei wird ein doppelsträngiger Oligonukleotidkomplex mit einer Primerbindungsstelle verwendet (Fig. 3a). Dieser wird mit kommerziell erhältlichen Ligasen an die DNA-Fragmente ligiert ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press). Es ist wichtig, dass nur eine einzige Primerbindungsstelle an das DNA-Fragment ligiert wird. Das erreicht man z.B. durch eine Modifikation einer Seite des Oligonukleotidkomplexes an beiden Strängen (Fig. 3b). Die Resultate nach der Ligation bzw. nach anschließender Denaturierung sind in Fig. 3c und 3d dargestellt. Die modifizierenden Gruppen am Oligonukleotidkomplex können zur Immobilisation dienen. Die Synthese und die Modifikation eines solchen Oligonukleotidkomplexes kann nach standardisierten Vorschriften durchgeführt werden. Zur Synthese kann z.B. der DNA-Synthesizer 380 A Applied Biosystems verwendet werden. Oligonucleotide mit einer bestimmten Zusammensetzung mit oder ohne Modifikationen sind aber auch als Auftragssynthese kommerziell erhältlich, z.B. von MWG-Biotech GmbH, Germany.

b) Nukleotid-Tailing:

Statt der Ligation mit einem Oligonukleotid kann man mit einer terminalen Deoxynucleotidyltransferase mehrere (z.B. zwischen

10 und 20) Nukleosid-monophosphate an das 3'-Ende eines ss-DNA-Fragments anknüpfen ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press, "Method in Enzymology" 1999 v.303, S.37-38) (Fig. 4), z.B. mehrere Guanosin-Monophosphate ((G)n-Tailing genannt). Das entstehende Fragment wird zur Bindung des Primers, in diesem Beispiel eines (C)n-Primers, verwendet.

#### 4.3.3 Einzelstrang-Vorbereitung

Für die Sequenzierungsreaktion werden einzelsträngige NSKFs benötigt. Falls das Ausgangsmaterial in doppelsträngiger Form vorliegt, gibt es mehrere Möglichkeiten, aus doppelsträngiger DNA eine einzelsträngige Form zu erzeugen (z.B. Hitze-Denaturierung oder Alkali-Denaturierung) ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press).

#### 4.3.4 Primer für die Sequenzierungsreaktion

Dieser hat die Funktion, den Start an einer einzigen Stelle des NSKF zu ermöglichen. Er bindet an die Primerbindungsstelle im NSKF. Die Zusammensetzung und die Länge des Primers sind nicht eingeschränkt. Außer der Startfunktion kann der Primer auch andere Funktionen übernehmen, wie z.B. eine Verbindung zur Reaktionsoberfläche zu schaffen. Primer sollten so an die Länge und Zusammensetzung der Primerbindungsstelle angepaßt werden, dass der Primer den Start der Sequenzierungsreaktion mit der jeweiligen Polymerase ermöglicht.

Bei der Verwendung unterschiedlicher, beispielsweise natürlich in der ursprünglichen Gesamtsequenz vorkommender Primerbindungsstellen, werden die für die jeweilige Primerbindungsstelle sequenzspezifischen Primer verwendet. In diesem Fall wird für die Sequenzierung ein Primergemisch eingesetzt.

Bei einer einheitlichen, beispielsweise durch die Ligation an



die NSKFs angekoppelten Primerbindungsstelle wird ein einheitlicher Primer verwendet.

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100 NTs, optimalerweise zwischen 15 und 30 NTs. Der Primer kann eine Funktionsgruppe tragen, die zur Immobilisierung des NSKF dient, beispielsweise ist eine solche Funktionsgruppe eine Biotingruppe (s. Abschnitt Immobilisierung). Sie soll die Sequenzierung nicht stören. Die Synthese eines solchen Primers kann z.B. mit dem DNA-Synthesizer 380 A Applied Biosystems ausgeführt werden oder aber als Auftragssynthese bei einem kommerziellen Anbieter, z.B. MWG-Biotech GmbH, Germany erstellt werden).

Der Primer kann vor der Hybridisierung an die zu analysierenden NSKFs auf der Oberfläche mit verschiedenen Techniken fixiert oder direkt auf der Oberfläche synthetisiert werden, beispielsweise nach (McGall et al. US Patent 5412087, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Fodor et al. Science 1991 v.285 S.767, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679).

Die Primer werden auf der Oberfläche beispielsweise in einer Dichte zwischen 10 bis 100 pro 100  $\mu\text{m}^2$ , 100 bis 10.000 pro 100  $\mu\text{m}^2$  oder 10.000 bis 1.000.000 pro 100  $\mu\text{m}^2$  gebunden.

Der Primer oder das Primergemisch wird mit NSKFs unter Hybridisierungsbedingungen inkubiert, die ihn selektiv an die Primerbindungsstelle des NSKF binden lassen. Diese Primer-Hybridisierung (Annealing) kann vor (1), während (2) oder nach (3) der Bindung der NSKFs an die Oberfläche erfolgen. Die Optimierung der Hybridisierungsbedingungen hängt von der genauen Struktur der Primerbindungsstelle und des Primers ab

und lässt sich nach Rychlik et al. NAR 1990 v.18 S.6409 berechnen. Im folgenden werden diese Hybridisierungsbedingungen als standardisierte Hybridisierungsbedingungen bezeichnet.

Falls eine für alle NSKFs gemeinsame Primerbindungsstelle mit bekannter Struktur beispielsweise durch Ligation eingeführt wird, können Primer mit einheitlicher Struktur eingesetzt werden. Die Primerbindungsstelle kann an ihrem 3'-Ende eine funktionelle Gruppe tragen, die z.B. zur Immobilisation dient. Beispielsweise ist diese Gruppe eine Biotin-Gruppe. Der Primer hat eine zur Primerbindungsstelle komplementäre Struktur.

Ein Beispiel einer Primerbindungsstelle und eines Primers ist nachfolgend dargestellt.

5'TAATACGACTCACTATAGG3'	Primer (T7-19-Primer)
Biotin-3'ATTATGCTGAGTGATATCC5'	Primerbindungsstelle

#### 4.3.5 Fixierung von NSKF-Primer-Komplexe an die Oberfläche (Bindung bzw. Immobilisierung von NSKFs).

Ziel der Fixierung (Immobilisierung) ist es, NSKF-Primer-Komplexe auf einer geeigneten planen Oberfläche in einer Art und Weise zu fixieren, dass eine zyklische enzymatische Sequenzierungsreaktion ablaufen kann. Dies kann beispielsweise durch Bindung des Primers (s.o.) oder des NSKF an die Oberfläche erfolgen.

Die Reihenfolge der Schritte bei der Fixierung von NSKF-Primer-Komplexen kann variabel sein:

- 1) Die NSKF-Primer-Komplexe können zunächst in einer Lösung durch Hybridisierung (Annealing) gebildet und anschließend an die Oberfläche gebunden werden.
- 2) Primer können zunächst auf einer Oberfläche gebunden werden und NSKFs anschließend an die gebundenen Primer hybridisiert werden, wobei NSKF-Primer-Komplexe entstehen (NSKFs indirekt an die Oberfläche gebunden)

- 3) Die NSKFs können zunächst an die Oberfläche gebunden werden (NSKFs direkt an die Oberfläche gebunden) und im anschließenden Schritt die Primer an die gebundenen NSKFs hybridisiert werden, wobei NSKF-Primer-Komplexe entstehen.

Die Immobilisierung der NSKFs an die Oberfläche kann daher durch direkte oder indirekte Bindung erfolgen.

Oberfläche und Reaktionsoberfläche sind vorliegend als gleichwertige Begriffe aufzufassen, außer wenn explizit auf eine andere Bedeutung hingewiesen wird. Als Reaktionsoberfläche dient die Oberfläche einer festen Phase eines beliebigen Materials. Dieses Material ist vorzugsweise enzymatischen Reaktionen gegenüber inert und verursacht keine Störungen der Detektion. Silicon, Glas, Keramik, Kunststoff (z.B. Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material, das diesen funktionellen Anforderungen genügt, kann verwendet werden. Vorzugsweise ist die Oberfläche nicht verformbar, denn sonst ist mit einer Verzerrung der Signale bei der wiederholten Detektion zu rechnen.

Falls eine gelartige feste Phase (Oberfläche eines Gels) verwendet wird, so kann dieses Gel z.B. ein Agarose- oder Polyacrylamidgel sein. Das Gel ist vorzugsweise für Moleküle mit einer Molekularmasse unter 5000 Da frei passierbar (beispielsweise kann ein 1 bis 2% Agarose-Gel oder 10 bis 15% Polyacrylamid Gel verwendet werden). Eine solche Geloberfläche hat anderen festen Oberflächen gegenüber den Vorteil, dass es zu einer wesentlich geringeren unspezifischen Bindung von NT\*s an die Oberfläche kommt. Durch die Bindung der NSKF-Primer-Komplexe auf der Oberfläche ist die Detektion der Fluoreszenzsignale von eingebauten NTs\* möglich. Die Signale von freien NTs\* werden nicht detektiert, weil sie nicht an das Material des Gels binden und somit nicht immobilisiert werden. Das Gel ist vorzugsweise auf einer festen Unterlage befestigt (Fig. 5a). Diese feste Unterlage kann Silicon, Glas, Keramik,

Kunststoff (z.B. Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material sein.

Die Dicke des Gels beträgt vorzugsweise nicht mehr als 0,1 mm. Die Geldicke ist vorzugsweise größer als die einfache Tiefenschärfe des Objektivs sein, damit unspezifisch an die feste Unterlage gebundene NTs nicht in die Fokusebene gelangen und damit detektiert werden. Wenn die Tiefenschärfe z.B. 0,3  $\mu\text{m}$  beträgt, so liegt die Geldicke vorzugsweise zwischen 1  $\mu\text{m}$  und 100  $\mu\text{m}$ . Die Oberfläche kann als eine kontinuierliche Oberfläche oder als diskontinuierliche, aus einzelnen kleinen Bestandteilen (z.B. Agarose-Kügelchen) zusammengesetzte Oberfläche hergestellt werden (Fig.5b). Die Reaktionsoberfläche muß groß genug sein, um die notwendige Anzahl der NSKFs bei entsprechender Dichte immobilisieren zu können. Die Reaktionsoberfläche sollte vorzugsweise nicht größer als 20  $\text{cm}^2$  sein.

Die verschiedenen Zyklusschritte erfordern einen Austausch der unterschiedlichen Reaktionslösungen über der Oberfläche. Die Reaktionsoberfläche ist vorzugsweise Bestandteil eines Reaktionsgefäßes. Das Reaktionsgefäß ist wiederum vorzugsweise Bestandteil einer Reaktionsapparatur mit Durchflußvorrichtung. Die Durchflußvorrichtung ermöglicht einen Austausch der Lösungen im Reaktionsgefäß. Der Austausch kann mit einer durch einen Computer gesteuerten Pumpvorrichtung oder manuell erfolgen. Wichtig dabei ist, dass die Oberfläche nicht austrocknet. Vorzugsweise beträgt das Volumen des Reaktionsgefäßes weniger als 50  $\mu\text{l}$ . Idealerweise beträgt sein Volumen weniger als 1  $\mu\text{l}$ . Ein Beispiel eines solchen Durchflußsystems ist in Fig.6 gegeben.

Falls die Fixierung der NSKF-Primer-Komplexe auf der Oberfläche über die NSKFs erfolgt, kann dies beispielsweise durch die Bindung der NSKFs an einem der beiden Ketten-Enden erfolgen. Dies kann durch entsprechende kovalente, affine oder andere Bindungen erreicht werden. Es sind viele Beispiele der

Immobilisierung von Nukleinsäuren bekannt (McGall et al. US Patent 5412087, Nikiforov et al. US Patent 5610287, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Rasmussen et al. Analytical Biochemistry v.198, S.138, Allemand et al. Biophysical Journal 1997, v.73, S.2064, Trabesinger et al. Analytical Chemistry 1999, v.71, S.279, Osborne et al. Analytical Chemistry 2000, v.72, S.3678, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679). Die Fixierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der NSKFs enthaltenden Probe auf der planen Oberfläche erreicht werden.

Die NSKFs werden auf der Oberfläche beispielsweise in einer Dichte zwischen 10 und 100 NSKFs pro  $100 \mu\text{m}^2$ , 100 bis 10.000 pro  $100 \mu\text{m}^2$ , 10.000 bis 1.000.000 pro  $100 \mu\text{m}^2$  gebunden.

Die für die Detektion notwendige Dichte von extensionsfähigen NSKF-Primer-Komplexen beträgt ca. 10 bis 100 pro  $100 \mu\text{m}^2$ . Sie kann vor, während oder nach der Hybridisierung der Primer an die Genprodukte erreicht werden.

Beispielhaft werden im folgenden einige Methoden zur Bindung von NSKF-Primer-Komplexen näher dargestellt: In einer Ausführungsform erfolgt die Immobilisierung der NSKFs über Biotin-Avidin oder Biotin-Streptavidin-Bindung. Dabei wird Avidin oder Streptavidin auf der Oberfläche kovalent gebunden, das 5'-Ende des Primers enthält Biotin. Nach der Hybridisierung der markierten Primer mit den NSKFs (in Lösung) werden diese auf der mit Avidin/Streptavidin beschichteten Oberfläche fixiert. Die Konzentration der mit Biotin markierten Hybridisierungs-Produkte sowie die Zeit der Inkubation dieser Lösung mit der Oberfläche wird so gewählt, dass eine für die Sequenzierung geeignete Dichte bereits in

diesem Schritt erreicht wird.

In einer anderen bevorzugten Ausführungsform werden die für die Sequenzierungsreaktion geeigneten Primer vor der Sequenzierungsreaktion auf der Oberfläche mit geeigneten Methoden fixiert (s.o.). Die einzelsträngigen NSKFs mit jeweils einer Primerbindungsstelle pro NSKF werden damit unter Hybridisierungsbedingungen inkubiert (Annealing). Dabei binden sie an die fixierten Primer und werden dadurch gebunden (indirekte Bindung), wobei Primer-NSKF-Komplexe entstehen. Die Konzentration der einzelsträngigen NSKFs und die Hybridisierungsbedingungen werden so gewählt, dass man eine für die Sequenzierung geeignete Immobilisationsdichte von 10 bis 100 extensionsfähigen NSKF-Primer-Komplexen pro  $100\ \mu\text{m}^2$  erreicht. Nach der Hybridisierung werden ungebundene NSKFs durch einen Waschschrift entfernt. Bei dieser Ausführungsform wird eine Oberfläche mit einer hohen Primerdichte bevorzugt, z.B. ca. 1.000.000 Primer pro  $100\ \mu\text{m}^2$  oder noch höher, da die gewünschte Dichte an NSKF-Primer-Komplexen schneller erreicht wird, wobei die NSKFs nur an einen Teil der Primer binden.

In einer anderen Ausführungsform werden die NSKFs an die Oberfläche direkt gebunden (s.o.) und anschließend mit Primern unter Hybridisierungsbedingungen inkubiert. Bei einer Dichte von ca. 10 bis 100 NSKFs pro  $100\ \mu\text{m}^2$  wird man versuchen alle verfügbaren NSKFs mit einem Primer zu versehen und für die Sequenzierungsreaktion verfügbar zu machen. Dies kann z.B. durch hohe Primerkonzentration, beispielsweise 1 bis 100 mmol/l, erreicht werden. Bei einer höheren Dichte der fixierten NSKFs auf der Oberfläche, beispielsweise 10.000 bis 1.000.000 pro  $100\ \mu\text{m}^2$ , kann die für die optische Detektion notwendige Dichte der NSKF-Primer-Komplexe während der Primer-Hybridisierung erreicht werden. Dabei sind die Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer, Primerkonzentration) so zu wählen, dass die Primer nur an einen Teil der immobilisierten NSKFs binden.

Falls die Oberfläche einer festen Phase (z.B. Silikon oder Glas) zur Immobilisation verwendet wird, wird vorzugsweise eine Blockierungslösung auf die Oberfläche vor dem Schritt (a) in jedem Zyklus gebracht, die zur Vermeidung einer unspezifischen Adsorption von NTs\* an der Oberfläche dient. Diese Bedingungen für eine Blockierungslösung erfüllt beispielsweise eine Albuminlösung (BSA) mit einem pH-Wert zwischen 8 und 10.

#### 4.4 Wahl der Polymerase

Als Polymerasen eignen sich prinzipiell alle DNA-abhängigen DNA-Polymerasen ohne 3'-5' Exonuklease-Aktivität (DNA-Replication" 1992 Ed. A.Kornberg, Freeman and company NY), z.B. modifizierte T7-Polymerase vom Typ "Sequenase Version 2" (Amersham Pharmacia Biotech), 3'-5' exonuklease freies Klenow Fragment der DNA-Polymerase I (Amersham Pharmacia Biotech), Polymerase Beta verschiedenen Ursprungs (Animal Cell DNA Polymerases" 1983, Fry M., CRC Press Inc., kommerziell erhältlich bei Chimex) thermostabile Polymerasen wie beispielsweise Taq-Polymerase (GibcoBRL), proHATM Polymerase (Eurogentech).

Polymerasen mit 3'-5' Exonuklease-Aktivität können eingesetzt werden (z.B. Klenow-Fragment der E.coli-Polymerase I), sofern Reaktionsbedingungen gewählt werden, die vorhandene 3'-5' Exonuklease-Aktivität unterdrücken, wie z.B. ein niedriger pH-Wert (pH 6.5) beim Klenow-Fragment (Lehman and Richardson, J. Biol. Chem. 1964 v.239 S.233) oder Zugabe von NaF zur Einbaureaktion. Eine andere Möglichkeit besteht in der Verwendung von NTs\* mit einer Phosphorothioate-Verbindung (Kunkel et al. PNAS 1981, v.78 S.6734). Dabei werden eingebaute NTs\* von der 3'-5' Exonuklease-Aktivität der Polymerase nicht angegriffen. Im folgenden werden all diese Polymerasearten als "Polymerase" bezeichnet.

## 4.5 Chemie

### 4.5.1 Allgemeine NT-Struktur

In den erfindungsgemäßen Verfahren können unterschiedliche NT\*s verwendet werden (vorzugsweise 2'-deoxy-Nukleotid-Triphosphate), die an ihrer 3'-Position des Riboseringes einen Substituenten tragen. Dieser Substituent kann alleine oder zusammen mit dem Fluoreszenzfarbstoff zur Termination der Einbaureaktion führen und kann unter milden Bedingungen vom Nukleotid abgespalten werden. An diesen Substituenten ist ein für das jeweilige NT\* charakteristischer Fluoreszenzfarbstoff angekoppelt, so dass der Substituent auch die Rolle eines Linkers zwischen dem Nukleotid und dem Fluoreszenzfarbstoff übernimmt. Der Fluoreszenzfarbstoff wird vorzugsweise an diesen Linker durch eine unter milden Bedingungen spaltbare Bindung angekoppelt.

Unter ~~den~~ milden Bedingungen werden Spaltungsbedingungen verstanden, die weder zur Denaturierung des Primer-Nukleinsäure-Komplexes führen, noch zur Spaltung seiner einzelner Bestandteile.

Formeln (1-3) stellen Beispiele für die reversiblen spaltbaren Terminatoren dar:

- 1) NT-3'-O-S(1)-F
- 2) NT-3'-O-S(2)-N-F
- 3) NT-3'-O-S(2)-N-L-F

NT-3'-O - stellt den 2'-Deoxy-Nukleosid-Triphosphat-Rest dar.

S(1) - stellt einen Substituenten (Formel 1) dar, der unter milden Bedingungen vom NT\* abgespalten werden kann. An diesen Substituenten ist ein Fluoreszenzfarbstoff (F) gekoppelt.

S(2)-N - stellt einen weiteren Substituenten (Formel 2 und



3) dar, der unter milden Bedingungen vom NT\* abgespalten werden kann. Dieser Substituent ist mit dem Fluoreszenzfarbstoff (F) durch eine unter milden Bedingungen spaltbare Gruppe (N) verbunden. Der Fluoreszenzfarbstoff kann unmittelbar an die spaltbare Gruppe (Formel 2) oder durch einen weiteren Linker (L) (Formel 3) gekoppelt sein.

Beispiele für NT\*-Strukturen, NT\*-Synthese, zur Polymerase-Wahl für die Einbaureaktion, Reaktionsbedingungen der NT\*-Einbaureaktion und Abspaltungsreaktion sind in (Kwiatkowski WO-Patent 01/25247, Kwiatkowski US-Patent 6.255.475, Conard et al. US-Patent 6.001.566, Dower (US Patent 5.547.839), Canard et al. (US Patent 5.798.210), Rasolonjatovo (Nucleosides & Nucleotides 1999, v.18 S.1021), Metzker et al. (NAR 1994, v.22, S.4259), Welch et al. (Nucleosides & Nucleotides 1999, v.18, S.197) beschrieben.

#### 4.5.2 Marker, Fluorophore

Jedes Nukleotid ist mit einem charakteristischen Marker (F) markiert. Der Marker ist ein fluoreszierender Farbstoff. Die Wahl ist nicht eingeschränkt, sofern der Farbstoff folgenden Anforderungen genügt:

a) Die verwendete Detektionsapparatur muß diesen Marker als einziges Molekül gebunden an DNA unter milden Bedingungen (vorzugsweise Reaktionsbedingungen) identifizieren können. Die Farbstoffe haben vorzugsweise große Photostabilität. Ihre Fluoreszenz wird vorzugsweise von der DNA nicht oder nur unwesentlich gequenchet.

b) Der an das NT gebundene Farbstoff darf keine irreversible Störung der enzymatischen Reaktion verursachen.

c) mit dem Farbstoff markierte NTs\* müssen von der Polymerase in die Nukleinsäurekette eingebaut werden.

d) Bei einer Markierung mit verschiedenen Farbstoffen sollen diese Farbstoffe keine beträchtlichen Überlappungen in ihren Emissionsspektren aufweisen.

Beispielsweise sind einige Fluorophore, die im Rahmen der vorliegenden Erfindung verwendbar sind, in "Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes mit Strukturformeln zusammengestellt. Erfindungsgemäß werden vorzugsweise folgende Farbstoffklassen als Marker eingesetzt: Cyanin-Farbstoffe und deren Abkömmlinge (z.B. Cy2, Cy3, Cy5, Cy7 Amersham Pharmacia Biotech, Waggoner US-Patent 5.268.486), Rhodamine und deren Abkömmlinge (z.B. TAMRA, TRITC, RG6, R110, ROX, Molecular Probes, s. Handbuch), Xanthene-Derivate (z.B. Alexa 568, Alexa 594, Molecular Probes, Mao et al. US-Patent 6.130.101) und Porphyrine (Porphyrin-Systems, Deutschland). Diese Farbstoffe sind kommerziell erhältlich.

Dabei kann man je nach spektralen Eigenschaften und vorhandener Apparatur entsprechende Farbstoffe auswählen. Die Farbstoffe werden an das NT\* über einen spaltbaren Linker gebunden. Die Farbstoffe können an den Linker z.B. über Thiocyanat- oder Ester-Bindung gekoppelt werden ("Handbook of Fluorescent Probes und Research Chemicals" 6th ed. 1996, R.Haugland, Molecular Probes, Jameson et al. Methods in Enzymology 1997 v.278 S.363, Waggoner Methods in Enzymology 1995 v.246 S.362)

#### 4.5.3 Spaltbare Bindung zwischen dem Nukleotid und dem Substituenten, Spaltung.

Der zur Termination führende Substituent ist an das NT durch eine unter milden Bedingungen spaltbare Bindung gekoppelt. Beispiele für diese Verbindungen stellen Ester und Acetale dar.

Die Spaltung der Ester erfolgt vorzugsweise im basischen pH-

Bereich (z.B. 9 bis 11). Die Spaltung von Acetalen erfolgt im saueren Bereich (z.B. zwischen 3 und 4).

Ester können auch enzymatisch durch Polymerasen oder Esterasen abgespalten werden.

In einer bevorzugten Ausführungsform der Erfindung wird der Substituent zusammen mit dem Fluoreszenzfarbstoff in einem Schritt abgespalten.

#### 4.5.4 Spaltbare Bindung zwischen dem Substituenten und dem Fluoreszenzfarbstoff, Spaltung.

In einer anderen bevorzugten Ausführungsform der Erfindung ist der Fluoreszenzfarbstoff an den Substituenten durch eine unter milden Bedingungen spaltbare Gruppe gekoppelt.

Vorzugsweise gehört die genannte Gruppe zu chemisch oder enzymatisch spaltbaren oder photolabilen Verbindungen.

Ester-, Thioester-, Disulfid-Verbindungen und photolabile Verbindungen eignen sich besonders gut als spaltbare Verbindung zwischen dem Substituenten und dem Fluoreszenzfarbstoff.

Als Beispiele von chemisch spaltbaren Gruppen sind Ester-, Thioester- und Disulfid-Verbindungen bevorzugt (Chemistry of protein conjugation and crosslinking`` Shan S. Wong 1993 CRC Press Inc., Herman et al. Method in Enzymology 1990 v.184 S.584, Lomant et al. J.Mol.Biol. 1976 v.104 243, "Chemistry of carboxylic acid and esters" S.Patai 1969 Interscience Publ.). Beispiele für photolabile Verbindungen können in folgenden Literaturstellen gefunden werden: "Protective groups in organic synthesis" 1991 John Willey & Sons, Inc., V. Pillai Synthesis 1980 S.1, V. Pillai Org.Photochem. 1987 v.9 S.225, Dissertation Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese`` H. Giegrich, 1996, Konstanz, Dissertation Neue photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese`` S.M.Bühler, 1999,

Konstanz).

Der Spaltungsschritt ist in jedem Zyklus vorhanden und muß unter milden Bedingungen verlaufen, so dass die Nukleinsäuren nicht beschädigt oder modifiziert werden.

Die Spaltung läuft bevorzugt chemisch (z.B. in milder saurer oder basischer Umgebung für eine Ester-Verbindung oder durch Zugabe eines Reduktionsmittels, z.B. Dithiothreitol oder Mercaptoethanol (Sigma) bei der Spaltung einer Disulfid-Verbindung), oder physikalisch (z.B. durch Beleuchtung der Oberfläche mit Licht einer bestimmten Wellenlänge für die Spaltung einer photolabilen Gruppe, Dissertation ~~Neue~~ photolabile Schutzgruppen für die lichtgesteuerte Oligonucleotidsynthese` H. Giegrich, 1996, Konstanz) ab.

In dieser Ausführungsform wird nach der Detektion zunächst der Fluoreszenzfarbstoff abgespalten und erst dann der an die 3'-Position gekoppelte, zur Termination führende Substituent.

#### 4.5.5 Farbiges Kodierungsschema, Anzahl der Farbstoffe

Jedes NT\* muß eindeutig mit einem charakteristischen Farbstoff markiert sein. Vorzugsweise kann man einen Zyklus durchführen mit:

- a) vier verschieden markierten NT\*s
- b) zwei verschieden markierten NT\*s
- c) einem markierten NT\*
- d) zwei verschieden markierten NT\*s und zwei unmarkierten NTs,

(auch andere Kombinationen sollten einem Fachmann naheliegend erscheinen)

d.h.

a) Man kann alle 4 NTs mit verschiedenen Farbstoffen markieren und alle 4 gleichzeitig in die Reaktion einsetzen. Dabei erreicht man die Sequenzierung einer Nukleinsäurekette

mit einer minimalen Anzahl von Zyklen. Diese Variante der Erfindung stellt allerdings hohe Anforderungen an das Detektionssystem: 4 verschiedene Farbstoffe müssen in jedem Zyklus identifiziert werden.

b) Zur Vereinfachung der Detektion kann eine Markierung mit zwei Farbstoffen gewählt werden. Dabei werden 2 Paare von NTs\* gebildet, die jeweils verschieden markiert sind, z.B. A und G tragen die Markierung "X", C und U tragen die Markierung "Y". In die Reaktion in einem Zyklus (n) werden 2 unterschiedlich markierte NTs\* gleichzeitig eingesetzt, z.B. C\* in Kombination mit A\*, und im darauffolgenden Zyklus (n+1) werden dann U\* und G\* zugegeben.

c) Man kann auch nur einen einzigen Farbstoff zur Markierung aller 4 NTs\* verwenden und pro Zyklus nur ein NT\* einsetzen.

d) In einer technisch vereinfachten Ausführungsform werden pro Zyklus zwei unterschiedlich markierte NT's eingesetzt und zwei unmarkierte NTs (sogen. 2NT\*s / 2NTs-Methode). Diese Ausführungsform kann verwendet werden, um Varianten (z.B. Mutationen, oder alternativ gespleißte Gene) einer bereits bekannten Sequenz zu ermitteln.

Unter Reaktionsbedingungen erfolgt der Einbau von NT's in die NSKFs vorzugsweise so, daß an mehr als 50% der an der Sequenzierungsreaktion beteiligten NSKFs in einem Zyklus ein markiertes NT\* eingebaut wird, vorzugsweise an mehr als 90%. Das hängt damit zusammen, daß an manchen Nukleinsäureketten die Reaktion sehr langsam abläuft. Ein Einbau der NTs\* an jeder komplementären Position in jedem Zyklus wird angestrebt, ist aber nicht erforderlich, weil nur die erfolgreichen Einbaureaktionen detektiert und ausgewertet werden; eine verzögerte Reaktion im Nachfolgenden Zyklus führt nicht zu einem Sequenzierungsfehler.

Vorzugsweise wird für alle NTs\* dieselbe Polymerase

verwendet. Es können aber auch verschiedene Polymerasen für verschiedene NTs eingesetzt werden.

#### 4.6 Detektionsapparatur

Einzelne Moleküle auf einer Oberfläche kann man mit verschiedenen Methoden untersuchen. Es sind mehrere Verfahren bekannt: z.B. AtomForce-Mikroskopie, Elektronen-Mikroskopie, Nahfeld-Fluoreszenz-Mikroskopie, Weitfeld-Fluoreszenz-Mikroskopie, TIR-Mikroskopie usw. (Science 1999 v.283 1667, Unger et al. BioTechniques 1999 v.27 S.1008, Ishijaima et al. Cell 1998 v.92 S.161, Dickson et al. Science 1996 v.274 S.966, Xie et al. Science 1994 v.265 S.361, Nie et al. Science 1994 v.266 S.1018, Betzig et al. Science 1993 v.262 S.1422).

Erfindungsgemäß werden Fluoreszenz-Signale einzelner in die Nukleinsäurekette eingebauter NTs vorzugsweise mit einem Weitfeld-Fluoreszenz-Mikroskop (Epifluoreszenz) oder einem Laser-Scanning-Mikroskop (Epifluoreszenz) oder einem TIRF-Mikroskop (Total Internal Reflection Fluorescence Microscope).

Es sind verschiedene Varianten der Konstruktion einer solchen Apparatur möglich (Weston et al. J.Chem.Phys. 1998 v.109 S.7474, Trabesinger et al. Anal. Chem. 1999 v.71 S.279, Adachi et al. Journal of Microscopy 1999 v.195 S.125, Unger et al. BioTechniques 1999 v.27 S.1008, Ishijaima et al. Cell 1998 v.92 S.161, Dickson et al. Science 1996 v.274 S.966, Tokunaga et al. Bichem.Biophys.Res.Com. 1997 v.235 S.47, "Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press ). Unterschiede in ihrem konkreten Aufbau ergeben sich aus der Variation ihrer Einzelteile. Die Vorrichtung für das Anregungslicht kann z.B. auf der Basis eines Lasers, einer Lampe oder von Dioden funktionieren. Für

die Detektionsvorrichtung können sowohl CCD-Kameras als auch PMT dienen. Andere Beispiele für technische Details siehe ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press). Es ist nicht die Aufgabe dieser Erfindung, alle möglichen technischen Varianten einer Detektionsvorrichtung aufzuzählen. Der prinzipielle Aufbau einer geeigneten Apparatur wird in einem Schema Fig. 8 erläutert. Sie besteht aus folgenden Elementen:

- Lichtquelle zur Anregung der Fluoreszenz (1)
- Lichtleitender Teil (2)
- Scantisch (3)
- Vorrichtung zur Selektion von Spektren (4)
- Detektionsvorrichtung (5)
- Computer mit Steuerungs- und Analysefunktionen (6)

Diese Elemente der Apparatur können kommerziell erworben werden (Mikroskop-Firmen: Zeiss, Leica, Nikon, Olympus).

Im folgenden soll beispielsweise eine für die Detektion einzelner Moleküle geeignete Kombination aus diesen Elementen vorgestellt werden:

Weitfeld-Fluoreszenz-Mikroskop Axioplan 2 (Zeiss) mit  
Quecksilberdampflampe  
Objektiv Planneofluar 100x, NA 1.4 (Zeiss)  
Kamera Photometrix oder AxioCam (Zeiss)  
Computer mit Software zur Steuerung und Analyse

Nachfolgend soll die Vorgehensweise bei der Detektion erläutert werden. Man beachte dabei die allgemeinen Regeln der Fluoreszenzmikroskopie ("Confocal Laser Scanning Microscopy"

1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press).

Die Detektion umfaßt folgende Phasen:

- 1) Vorbereitung zur Detektion
- 2) Durchführung eines Detektionsschrittes in jedem Zyklus, wobei jeder Detektionsschritt als Scanvorgang abläuft und folgende Operationen umfaßt:
  - a) Einstellung der Position des Objektivs (X,Y-Achse),
  - b) Einstellung der Fokusebene (Z-Achse),
  - c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT' und Zuordnung des Signals zum jeweiligen NSKF,
  - d) Verschiebung zur nächsten Position auf der Oberfläche.

Die Signale von in die NSKFs eingebauten NTs\* werden durch das Abscannen der Oberfläche registriert. Der Scanvorgang kann in verschiedener Weise ausgeführt werden ("Confocal Laser Scanning Microscopy" 1997 Ed. Sheppard, BIOS Scientific Publishers, "New Techniques of optical microscopy and microspectroscopy" 1991 Ed. R.Cherry CRC Press, Inc., "Fluorescence microscopy" 1998 2.ed. Herman BIOS Scientific Publishers, "Handbook of biological confocal microscopy" 1995 J.Pawley Plenum Press). Beispielsweise wird ein diskontinuierlicher Scanvorgang gewählt. Dabei wird das Objektiv schrittweise über die Oberfläche bewegt (Fig. 8a), so dass von jeder Oberflächenposition ein zweidimensionales Bild (2D-Bild) entsteht (Fig 8b,c).

Dieses 2D-Bild kann mit verschiedenen Methoden erstellt werden: z.B. durch den Laser-Scan einer Position des Mikroskopfeldes (Laser-Scanning-Mikroskopie) oder durch eine



Kameraaufnahme an einer Position (vgl. Handbücher der Mikroskopie). Als Beispiel wird die Detektion einzelner Moleküle mit einer CCD-Kamera beschrieben.

Die Detektion wird schematisch am Beispiel der Sequenzierung eines 1Mb langen DNA-Stücks erläutert:

1) Vorbereitung zur Detektion:

Am Anfang wird festgelegt, wie viele NSKF-Sequenzen zur Rekonstruktion der ursprünglichen Sequenz analysiert werden müssen. Im Fall einer Rekonstruktion nach dem Schrotschuß-Verfahren ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257) spielen folgende Faktoren eine Rolle: 1) Von jedem NSKF wird bei der Sequenzierung eine Sequenz von ca. 300-500 NTs bestimmt. 2) Die Gesamtlänge der zu analysierenden Sequenz ist wichtig. 3) Bei der Sequenzierung muß ein bestimmtes Maß an Redundanz erreicht werden, um die Genauigkeit zu steigern und eventuelle Fehler zu korrigieren. Insgesamt ist zur Rekonstruktion des größten Teils der ursprünglichen Sequenz die etwa 10- bis 100-fache Menge an Rohsequenzen erforderlich, d.h. bei diesem Beispiel mit einer Mb, werden 10 bis 100 Mb Rohsequenzdaten gebraucht. Bei einer durchschnittlichen Sequenzlänge von 400 bp pro NSKF benötigt man entsprechend 25.000 bis 250.000 DNA-Fragmente.

2) Durchführung eines Detektionsschrittes in jedem Zyklus

Zur Sequenzierung müssen die Positionen der NSKFs bestimmt werden, damit man eine Grundlage für die Zuordnung der Signale hat. Die Kenntnis dieser Positionen erlaubt eine Aussage darüber, ob die Signale einzelner Moleküle von eingebauten NTs\* stammen oder von zufällig an die Oberfläche gebundenen NTs\*. Diese Positionen können mit verschiedenen Methoden identifiziert werden.

In einer bevorzugten Ausführungsform werden die Positionen gebundener NSKF-Primer-Komplexe während der Sequenzierung identifiziert. Dabei wird die Tatsache genutzt, dass die Signale von den in die Nukleinsäurekette eingebauten NTs<sup>\*</sup> immer dieselben Koordinaten haben. Das ist durch die Fixierung der Nukleinsäureketten gewährleistet. Die unspezifisch gebundenen NTs<sup>\*</sup> binden zufällig an verschiedenen Stellen der Oberfläche.

Zur Identifizierung der Positionen von fixierten NSKFs werden die Signale auf Übereinstimmung ihrer Koordinaten aus mehreren aufeinander folgenden Zyklen überprüft. Das kann z.B. am Anfang der Sequenzierung erfolgen. Die übereinstimmende Koordinaten werden als Koordinaten der DNA-Fragmente bewertet und gespeichert.

Das Scan-System muß reproduzierbar über mehrere Zyklen die Oberfläche abscannen können. X,Y und Z-Achsen-Einstellungen an jeder Oberflächenposition können von einem Computer kontrolliert werden. Stabilität und Reproduzierbarkeit der Einstellung von Objektivpositionen in jedem Scanvorgang entscheiden über die Qualität der Detektion und somit über die Identifizierung der Signale einzelner Moleküle.

#### a) Einstellung der Position des Objektivs (X,Y-Achse)

Die mechanische Instabilität der kommerziell erhältlichen Scantische und die geringe Reproduzierbarkeit der wiederholten Einstellung derselben X,Y-Positionen machen eine präzise Analysen der Signale einzelner Moleküle über mehrere Zyklen schwierig. Es existieren viele Möglichkeiten, eine Übereinstimmung der Koordinaten bei wiederholten Einstellungen zu verbessern bzw. mögliche Abweichungen zu kontrollieren. Als Beispiel wird eine Kontrollmöglichkeit angeführt. Nach einer groben mechanischen Einstellung der Objektivposition wird ein Kontrollbild von einem mit der Oberfläche fest verbundenen Muster gemacht. Auch wenn die mechanische Einstellung nicht

exakt dieselben Koordinaten aufweist (Abweichungen bis zu 10  $\mu\text{m}$  sind durchaus möglich), kann man mittels optischer Kontrolle eine Korrektur vornehmen. Das Kontrollbild vom Muster dient als Koordinatensystem für das Bild mit Signalen von eingebauten NTs. Eine Voraussetzung für eine solche Korrektur ist, dass keine weiteren Bewegungen der Oberfläche zwischen diesen beiden Aufnahmen gemacht werden. Signale von einzelnen Molekülen werden in Relation zum Muster gesetzt, so dass eine X,Y-Abweichung in der Musterposition gleiche X,Y-Abweichung in der Position der Signale einzelner Moleküle bedeutet. Das Kontrollbild vom Muster kann vor, während oder nach der Detektion einzelner Moleküle gemacht werden. Ein solches Kontrollbild muß entsprechend bei jeder Einstellung auf einer neuen Oberflächenposition gemacht werden.

#### b) Einstellung der Fokusebene (Z-Achse)

Die Oberfläche ist nicht absolut plan und weist verschiedene Unebenheiten auf. Dadurch verändert sich der Oberfläche-Objektiv-Abstand beim abscannen benachbarter Stellen. Diese Unterschiede im Abstand können dazu führen, dass einzelne Moleküle die Fokusebene verlassen und so der Detektion entgehen.

Aus diesem Grund ist es wichtig, dass beim Abscannen der Oberfläche eine reproduzierbare Einstellung der Fokusebene an jeder Objektivposition erreicht wird.

Es gibt verschiedene Möglichkeiten, die Fokusebene reproduzierbar einzustellen. Beispielsweise kann folgende Methode angewendet werden: Da die Anregung einzelner Moleküle zum Auslöschen ihrer Fluoreszenz führen kann, wird auf die Oberfläche ein Marker aufgebracht, der zur Einstellung der Fokusebene dient. Danach erfolgt die Detektion der Signale einzelner Moleküle. Der Marker kann beliebiger Natur sein (z.B. Farbstoff oder Muster), darf aber die Detektion und die Reaktion nicht beeinträchtigen.

c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT' und Zuordnung des Signals zum jeweiligen NSKF.

Das mit Hilfe des Detektionssystems erzeugte zweidimensionale Bild der Reaktionsoberfläche enthält die Signalinformationen von in die NSKFs eingebauten NT's. Diese müssen vor der weiteren Verarbeitung aus der Gesamtdatenmenge der Bildinformationen mit geeigneten Methoden extrahiert werden. Die dazu notwendigen Algorithmen zur Skalierung, Transformation und Filterung der Bildinformationen zählen zum Standardrepertoire der digitalen Bildverarbeitung und Mustererkennung (Haberäcker P. "Praxis der Digitalen Bildverarbeitung und Mustererkennung". Hanser-Verlag, München, Wien, 1995; Galbiati L.J. "Machine vision and digital image processing fundamentals". Prentice Hall, Englewood Cliffs, New Jersey, 1990). Die Signalextraktion erfolgt vorzugsweise über ein Grauwertbild, das die Helligkeitsverteilung der Reaktionsoberfläche für den jeweiligen Fluoreszenzkanal abbildet. Wenn bei der Sequenzierungsreaktion mehrere Nukleotide mit unterschiedlichen Fluoreszenz-Farbstoffen verwendet werden, kann zunächst für jedes verwendete fluoreszenzmarkierte Nukleotid (A,T,C,G oder U) ein separates Grauwert-Bild erzeugt werden. Dafür können prinzipiell 2 Verfahren angewendet werden:

1. Durch Verwendung von geeigneten Filtern (Zeiss-Filtersätze) wird für jeden Fluoreszenzkanal ein Grauwertbild erzeugt.

2. Aus einem aufgenommenen Mehrkanal-Farb-Bild werden mit Hilfe eines geeigneten Algorithmus durch ein Bildverarbeitungsprogramm die relevanten Farbkanäle extrahiert und jeweils als Grauwertbild einzeln weiterverarbeitet. Zur Kanalextraktion wird dabei ein für den jeweiligen Kanal spezifischer Farb-Schwellwertalgorithmus eingesetzt. So entstehen zunächst aus einem Mehrkanal-Farbbild einzelne Grauwertbilder 1 bis N. Diese Bilder definieren sich wie folgt:

$GB_n = (s(x,y))$  einkanalisches Grauwertbild  
 $N = \{1, \dots, \text{Anzahl der Fluoreszenzkanäle}\}$ .  
 $M = \{0, 1, \dots, 255\}$  Grauwertmenge  
 $S = (s(x,y))$  Bildmatrix des Grauwertbildes  
 $x = 0, 1, \dots, L-1$  Bildzeilen  
 $y = 0, 1, \dots, R-1$  Bildspalten  
 $(x,y)$  Ortskoordinaten eines Bildpunktes  
 $s(x,y) \in M$  Grauwert des Bildpunktes.

Aus dieser Datenmenge wird nun durch ein geeignetes Programm die relevante Bildinformation extrahiert. Ein solches Programm sollte folgende Arbeitsschritte realisieren:

Für  $GB_1$  bis  $GB_N$  durchführen:

I. Vorverarbeitung des Bildes, so zum Beispiel gegebenenfalls Reduktion des durch die Digitalisierung der Bildinformation entstandenen Bildrauschens, etwa durch Grauwertglättung.

II. Prüfung jedes Bildpunkt  $(x,y)$  des Grauwertbildes, ob dieser Punkt im Zusammenhang mit den ihn umgebenden unmittelbaren und weiter entfernten Nachbarbildpunkten die Eigenschaften eines Fluoreszenzpunktes erfüllt. Diese Eigenschaften hängen unter anderem von der verwendeten Detektionsapparatur und der Auflösung des Grauwertbildes ab. Sie können beispielsweise ein typisches Verteilungsmuster von Helligkeits-Intensitätswerten über einer den Bildpunkt umgebenden Matrix darstellen. Die dazu verwendbaren Methoden der Bildsegmentierung reichen von einfachen Schwellwertverfahren bis hin zur Verwendung neuronaler Netze.

Erfüllt ein Bildpunkt  $(x,y)$  diese Anforderungen, dann folgt ein Vergleich mit den Koordinaten von in bisher durchgeführten Sequenzierungszyklen identifizierten NSKFs. Bei einer Übereinstimmung erfolgt die Zuordnung des Signals mit dem aus dem jeweiligen Fluoreszenzkanal hervorgehenden Nukleotid zu diesem

NSKF. Signale mit nicht übereinstimmenden Koordinaten werden als Hintergrundsignale bewertet und verworfen. Die Analyse der Signale kann parallel zum Scanvorgang erfolgen.

In einer beispielhaften Ausführung wurde ein 8-Bit-Grauwertbild mit einer Auflösung von 1317 x 1035 Pixel verwendet. Um die durch die Digitalisierung entstandenen Veränderungen am Bild zu reduzieren, erfolgte zunächst eine Vorverarbeitung des Gesamtbildes: Jedem Bildpunkt wurde der Mittelwert der Helligkeiten seiner 8-Nachbarn zugewiesen. Bei der gewählten Auflösung entsteht dadurch ein für einen Fluoreszenzpunkt typisches Muster eines zentralen Bildpunkt mit dem größten Helligkeitswert und Nachbarbildpunkten mit nach allen Seiten hin abfallenden Helligkeiten. Erfüllte ein Bildpunkt diese Kriterien und überschritt der zentrifugale Helligkeitsabfall einen bestimmten Schwellenwert (zur Exklusion zu schwacher Fluoreszenzpunkte), dann wurde dieser zentrale Bildpunkt als Koordinate eines Fluoreszenzpunktes gewertet.

d) Verschiebung des Objektivs zur nächsten Position auf der Oberfläche. Nach der Detektion der Signale einzelner Moleküle wird das Objektiv über einer anderen Position der Oberfläche positioniert.

Insgesamt kann beispielsweise eine Folge von Aufnahmen mit der Kontrolle der X,Y-Position, der Einstellung der Fokusebene und mit der Detektion einzelner Moleküle bei jeder neuen Objektivposition gemacht werden. Diese Schritte können durch einen Computer gesteuert werden.

#### 4.7 Zeitlicher Ablauf der Verfahrensschritte

Der Scanvorgang sowie die biochemische Reaktion nehmen eine gewisse Zeit in Anspruch. Wenn man diese Vorgänge nacheinander schaltet, kann man eine optimale Leistung der Apparatur erreichen. In einer bevorzugten Ausführung wird die Reaktion auf

zwei getrennten Oberflächen durchgeführt.

Als Beispiel kann eine Oberfläche mit gebundenen NSKF-Primer-Komplexen in 2 räumlich isolierte Teile getrennt werden, so dass Reaktionen auf diesen beiden Teilen unabhängig voneinander ablaufen können. In einem anderen Beispiel können NSKFs auch von vornherein auf 2 getrennten Oberflächen immobilisiert werden.

Danach wird die Reaktion gestartet. Das Prinzip dabei ist, dass während auf einem Teil der Oberfläche die Reaktions- und Waschschrirte ablaufen, der zweite Teil abgescannt wird. Dadurch kann man einen kontinuierlichen Ablauf der Analyse erreichen und die Geschwindigkeit der Sequenzierung steigern.

Die Anzahl der Oberflächen, auf denen die Reaktion abläuft, kann auch größer als 2 sein. Das erscheint dann sinnvoll, wenn die Reaktion als zeitlich limitierender Schritt auftritt, d.h. die Detektion der Signale auf der Oberfläche schneller als die Reaktions- und Waschschrirte abläuft. Um die Gesamtdauer der Reaktion an die Detektionsdauer anzupassen, kann jeder einzelne Schritt der Reaktion auf einer einzelnen Oberfläche mit einer zeitlichen Verzögerung im Vergleich zur nächsten Oberfläche ablaufen.

Die Erfindung wird nachfolgend anhand von Beispielen verdeutlicht.

### Beispiele

#### Beispiel 1:

Sequenzanalyse mit 4 markierten NTs\*

Bei einer bevorzugten Ausführungsform der Erfindung werden alle vier in die Reaktion eingesetzten NTs\* mit Fluoreszenzfarbstoffen markiert.

1A. Rekonstruktion der ursprünglichen Sequenzen nach dem Schrotschuß-Prinzip ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). (Dieses Prinzip ist insbesondere bei der Analyse neuer, unbekannter Sequenzen geeignet.)

#### 1A-1Sequenzierung eines langen DNA-Stücks

Im folgenden soll anhand der Sequenzierung eines 1Mb langen DNA-Stückes schematisch die Sequenzierung langer Nukleinsäureketten dargestellt werden (Fig. 1). Der Sequenzierung liegt das Shotgun-Prinzip zugrunde ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257). Das zu analysierende Material wird für die Sequenzierungsreaktion vorbereitet, indem es in Fragmente von vorzugsweise 50 bis 1000 bp Länge zerlegt wird. Jedes Fragment wird anschließend mit einer Primerbindungsstelle und einem Primer versehen. Dieses Gemisch aus verschiedenen DNA-Fragmenten wird nun auf einer planen Oberfläche fixiert. Die nicht gebundenen DNA-Fragmente werden durch einen Waschschrift entfernt. Danach wird die Sequenzierungsreaktion an der gesamten Reaktions-



oberfläche durchgeführt. Zur Rekonstruktion einer 1 Mb langen DNA-Sequenz sollten die Sequenzen von NSKFs vorzugsweise länger als 300 NTs sein, durchschnittlich ca. 400 bp. Da pro Zyklus nur jeweils ein markiertes NT<sup>\*</sup> eingebaut wird, sind mindestens 400 Zyklen zur Sequenzierung notwendig.

Insgesamt ist zur Rekonstruktion der ursprünglichen Sequenz die etwa 10- bis 100-fache Menge an Rohsequenzen erforderlich, d.h. 10 bis 100 Mb. Bei einer durchschnittlichen Sequenzlänge von ca. 400 bp pro NSKF benötigt man entsprechend 25.000 bis 250.000 DNA-Fragmente, um mehr als 99,995% der Gesamtsequenz abzudecken.

Die ermittelten NSKF-Sequenzen stellen eine Population von überlappenden Teilsequenzen dar, die sich mit kommerziell erhältlichen Programmen zur Gesamtsequenz der NSK zusammenfügen lassen ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257).

#### 1A-2 Sequenzierung der Genprodukte am Beispiel der cDNA-Sequenzierung

In einer bevorzugten Ausführungsform können statt einer Sequenz mehrere Sequenzen in einem Ansatz analysiert werden. Die ursprünglichen Sequenzen können aus den gewonnen Rohdaten z.B. nach dem Schrotschuß-Prinzip rekonstruiert werden.

Zunächst werden NSKFs erzeugt. Man kann z.B. mRNA in eine doppelsträngige cDNA überführen und diese cDNA mit Ultraschall fragmentieren. Anschließend werden diese NSKFs mit einer Primerbindungsstelle versehen, denaturiert,

immobilisiert und mit einem Primer hybridisiert. Zu beachten ist bei dieser Variante der Probenvorbereitung, dass die cDNA-Moleküle unvollständige mRNA-Sequenzen darstellen können (Method in Enzymology 1999, v.303, S.19 und andere Artikel in diesem Band, "cDNA library protocols" 1997 Humana Press).

Eine andere Möglichkeit bei der Generierung einzelsträngiger NSKFs von mRNA besteht in der reversen Transkription der mRNA mit randomisierten Primern. Dabei werden viele relativ kurze antisense DNA-Fragmente gebildet (Zhang-J et al. Biochem.J. 1999 v.337 S.231, Ledbetter et al. J.Biol.Chem. 1994 v.269 S.31544, Kolls et al. Anal.Biochem. 1993 v.208 S.264, Decraene et al. Biotechniques 1999 v.27 S.962). Diese Fragmente können anschließend mit einer Primerbindungstelle versehen werden (s.o). Weitere Schritte entsprechen oben beschriebenen Vorgängen. Mit dieser Methode können komplette mRNA-Sequenzen (vom 5'- bis zum 3'-Ende) analysiert werden, da die randomisierten Primer über die gesamte Länge der mRNA binden.

Immobilisierte NSKFs werden mit einer der oben angeführten Ausführungsformen der Sequenzierung analysiert. Da mRNA-Sequenzen wesentlich weniger repetitive Sequenzen aufweisen als z.B. genomische DNA, kann die Anzahl der detektierten Signale der eingebauten NTs' von einem NSKF geringer als 300 sein und liegt vorzugsweise zwischen 20 und 1000. Die Anzahl der NSKFs, die analysiert werden müssen, errechnet sich nach denselben Prinzipien wie bei einer Schrotschuß-Rekonstruktion einer langen Sequenz.

Aus NSKF-Sequenzen werden nach den Prinzipien des Schrotschuß-Verfahrens die ursprünglichen Gensequenzen rekonstruiert.

Diese Methode erlaubt die gleichzeitige Sequenzierung von

vielen mRNAs ohne vorherige Klonierung.

#### 1B. Analyse von Sequenzvarianten

Die Bestätigung einer bereits bekannten Sequenz oder der Nachweis von Varianten dieser Sequenz stellt sehr viel geringere Ansprüche an die Länge und Redundanz der ermittelten NSKF-Sequenzen. Auch die Sequenzbearbeitung ist in diesem Fall einfacher. Die Vollsequenz braucht nicht neu rekonstruiert zu werden. Die NSKF-Sequenzen werden vielmehr mit Hilfe eines kommerziell erhältlichen Programms der Vollsequenz zugeordnet und eventuelle Abweichungen detektiert. Einem solchen Programm kann z.B. BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Die zu analysierende Sequenz wird mit einer der oben genannten Methoden in NSKFs überführt. Diese NSKFs werden mit dem erfindungsgemäßen Verfahren sequenziert, wobei man sowohl einen einheitlichen Primer und eine einheitliche Primerbindungsstelle als auch unterschiedliche, sequenzspezifische Primer und natürliche, in der zu untersuchenden Gesamtsequenz vorkommende Primerbindungsstellen verwenden kann. Anschließend werden die ermittelten Sequenzen von NSKFs nicht nach dem Schrotschuß-Verfahren zusammengestellt, sondern mit der Referenzsequenz verglichen und auf diese Weise ihren Positionen in der Vollsequenz zugeordnet. Dabei kann es sich um genomische oder cDNA-Sequenzen handeln.

Im Gegensatz zu einer Rekonstruktion nach dem Schrotschuß-Verfahren braucht man für die Analyse einer Sequenzvariante erheblich weniger Rohsequenzdaten. So kann die 5- bis 10- fache Rohsequenzmenge ausreichend für die Wiederherstellung einer neuen Variante einer Vollsequenz sein. Mit dem Schrotschuß-Verfahren wird für eine

Wiederherstellung eine 10- bis 100-fache Menge an Rohsequenzen benötigt ("Automated DNA sequencing and analysis" S. 231 ff. 1994 M. Adams et al. Academic Press, Huang et al. Genom Res. 1999 v.9 S.868, Huang Genomics 1996 v.33 S.21, Bonfield et al. NAR 1995 v.23 S.4992, Miller et al. J.Comput.Biol. 1994 v.1 S.257).

Die Länge der ermittelten NSKF-Sequenzen soll für eine eindeutige Zuordnung zu einer bestimmten Position in der Referenzsequenz ausreichend sein, so können z.B bereits Sequenzen mit einer Länge von 20 NTs (z.B. aus nicht repetitiven Abschnitten im menschlichen Genom) eindeutig identifiziert werden. Für die Vergleichsanalyse der repetitiven Abschnitte werden längere Sequenzen benötigt. Die genaue Länge der Sequenzen hängt dabei von der Aufgabenstellung ab. Vorzugsweise beträgt die Länge der ermittelten NSKF-Sequenzen bei der Analyse von nicht repetitiven Abschnitten mehr als 20 NTs. Für die Analyse der repetitiven Abschnitte liegt sie vorzugsweise über 500 NTs.

Die Zielsetzungen bei der Sequenzierung neuer Varianten einer bereits bekannten Vollsequenz können sehr unterschiedlich sein. Meist wird ein Vergleich der neu ermittelten Sequenz mit der bekannten Vollsequenz/Referenzsequenz angestrebt. Dabei können die beiden Sequenzen aus evolutionär unterschiedlich weit auseinanderliegenden Spezies stammen. Verschiedene Parameter der Zusammensetzung dieser beiden Sequenzen können verglichen werden. Als Beispiele für eine solche Analyse dienen: Mutations- oder Polymorphismusanalysen und die Analyse von alternativ gespleißten Genprodukten.

Nachfolgend soll schematisch und beispielhaft ein Vergleich der zu untersuchenden Sequenz mit einer Referenzsequenz ohne vorherige Rekonstruktion der zu analysierenden Sequenz betrachtet werden. Ein solcher

Vergleich kann z.B. zur Mutations- oder SNP-Analyse dienen.

#### 1B-1

Eine lange, zu analysierende Sequenz, z.B. 1 Mb, wird in NSKFs mit einer der oben genannten Methode geteilt. Diese NSKFs werden unter Verwendung einheitlicher Primer mit dem erfindungsgemäßen Verfahren sequenziert. Die ermittelten Sequenzen von jedem einzelnen NSKF werden direkt mit der Referenzsequenz verglichen. Die Referenzsequenz dient dabei als Grundlage für die Zuordnung ermittelter NSKF-Sequenzen, so dass die aufwendige Rekonstruktion nach dem Schrotschuß-Verfahren entfällt. Vorzugsweise beträgt die Länge der ermittelten NSKF-Sequenzen bei der Analyse von nicht-repetitiven Abschnitten mehr als 20 NTs. Für die Analyse der repetitiven Abschnitte liegt sie vorzugsweise über 500 NTs. Die Anzahl der zu analysierenden NSKFs richtet sich dabei nach der Gesamtlänge der zu untersuchenden Sequenz, der durchschnittlichen Länge der NSKF-Sequenzen und der notwendigen Genauigkeit der Sequenzierung. Bei einer durchschnittlichen Länge der ermittelten NSKF-Sequenz von 100 NTs, einer Gesamtlänge der zu untersuchenden Sequenz von 1 Mb und einer Genauigkeit, die der Rohsequenzermittlung entspricht (d.h. jede Stelle soll möglichst nur einmal sequenziert werden) benötigt man z.B. die ca. 5-fache Menge an Rohsequenzen, d.h. 5 Mb, weil die Verteilung der NSKFs über die Gesamtsequenz zufällig erfolgt. Insgesamt müssen 50.000 NSKFs analysiert werden, um mehr als 99% der Gesamtstrecke abzudecken.

Anschließend werden die ermittelten NSKF-Sequenzen mit Hilfe eines kommerziell erhältlichen Programms der Vollsequenz zugeordnet und eventuelle Abweichungen detektiert. Einem solchen Programm kann z.B. BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to

computational Biology" 1995 M.S. Waterman Chapman & Hall).

Beispiel 2:

Sequenzanalyse mit 2 markierten NTs' und 2 unmarkierten NTs (2NTs' / 2NTs-Methode).

In einer anderen Ausführungsform werden für die Analyse der Sequenzen 2 modifizierte NTs' und 2 unmodifizierte NTs eingesetzt.

Diese Methode eignet sich besonders zur Analyse der Sequenzvarianten (z.B. SNP- oder Mutationsanalyse) und setzt die Kenntnis einer Referenzsequenz voraus. Dabei wird die Vollsequenz nicht rekonstruiert, sondern die ermittelten Sequenzen werden mit Hilfe eines Programms der Referenzsequenz zugeordnet und eventuelle Abweichungen registriert. Einem solchen Programm kann z.B. der BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Diese Ausführungsform beruht auf dem Prinzip, dass eine Abfolge aus 2 Signalen (markierte NT's) genügend Informationen zur Identifizierung einer Sequenz enthalten kann. Die ermittelte Sequenz wird mit der Referenzsequenz verglichen und einer bestimmten Position zugeordnet, z.B.:

ACCAAAACACCC - ermittelte Sequenz (dCTP' und dATP' sind markiert)

ATCATCGTTTCGAAATATCGATCGCCTGATGCC - Referenzsequenz

A-C---C-AAA-A-C-A-C-CC (zugeordnete ermittelte Sequenz)  
ATCATCGTTTCGAAATATCGATCGCCTGATGCC (Referenzsequenz)

Die unbekannte, zu analysierende Variante der Referenzsequenz wird wie oben beschrieben zur Sequenzierung vorbereitet (NSK wird in NSKFs überführt, diese werden mit PBS ligiert, anschließend mit einem Primer hybridisiert und auf Reaktionsoberfläche immobilisiert). Auf diese Weise vorbereitete NSKFs werden mit 2NTs\*/2NTs-Methode sequenziert. Man erhält NSKF-Sequenzen, wobei jede NSKF-Sequenz eine Abfolge aus 2NTs\* darstellt. Um eine eindeutige Zuordnung der ermittelten Sequenz zu einer bekannten Referenzsequenz zu ermöglichen, muß diese Abfolge lang genug sein. Vorzugsweise beträgt die Länge der ermittelten NSKF-Sequenzen mehr als 40 NT's. Da 2 markierte NTs\* nur einen Teil der Sequenz darstellen, ist die Gesamtlänge des synthetisierten komplementären Strangs ca. doppelt so lang, wie die Abfolge der detektierten NTs\* (bei 40 detektierten NTs\* beträgt die Gesamtlänge z.B. durchschnittlich 80 NTs).

Zur Synthese eines komplementären Stranges werden 4 Nukleotide benötigt. Da die mit einem Fluoreszenzfarbstoff markierten NTs\* in der vorliegenden Erfindung als Semiterminatoren auftreten, d.h. die Termination ausschließlich bei Verfügbarkeit modifizierter NTs\* auftritt, müssen unmodifizierte NTs in einem zusätzlichen Schritt in jedem Zyklus in die Reaktion zugegeben werden. Die genaue Position dieses Schrittes in dem Zyklus kann variieren. Wichtig dabei ist, dass die markierten NTs\* und die unmodifizierte NTs getrennt verwendet werden.

Ein Zyklus bei dieser Ausführungsform kann beispielhaft folgendermaßen aussehen:

- a) Zugabe einer Lösung mit modifizierten NTs\* und Polymerasen auf die Oberfläche mit den bereitgestellten NSKFs
- b) Inkubation der immobilisierten Nukleinsäureketten mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind
- c) Waschen

- d) Detektion der Signale von einzelnen, modifizierten und in die den NSKFs komplementären neusynthetisierten Strängen eingebauten NTs'-Molekülen
- e) Entfernung der Markierung und der terminierenden Gruppe bei den eingebauten Nukleotiden
- f) Waschen
- g) Zugabe von 2 unmodifizierten NTs und Polymerasen
- h) Waschen.

Diese 2NT's/2NTs-Methode eignet sich beispielsweise für die SNP-Analyse einer genomischen Strecke eines Gens oder für doppelsträngige cDNA-Analyse. Ihr liegen folgende Prinzipien zugrunde:

- 1) Die genetische Information in jedem der beiden komplementären DNA-Stränge ist identisch, so dass fehlende Informationen in einem Strang durch die Information aus dem anderen Strang vervollständigt werden können.
- 2) Durch bestimmte Paarkombinationen markierter NTs' kann man mit nur 2 NTs' die komplette Information aus einer doppelsträngigen DNA erhalten. Zulässige Kombinationen von NT's bei dieser Ausführungsform sind: A'C'; A'G'; C'T'/C'U'; G'T'/G'U'. Bevorzugt wird die Kombination C' und U'.
- 3) Als Grundlage der Analyse dient eine bereits bekannte Referenzsequenz.
- 4) Die NSKFs stammen von beiden Strängen der zu analysierenden NSK und die ermittelten NSKF-Sequenzen decken die gesamte Länge der zu analysierenden Sequenz ab.

Am folgenden Beispiel wird erklärt, wie die Information aus einem doppelsträngigen DNA-Fragment mit nur 2 markierten NTs' gewonnen wird und wie die Unterschiede zur ursprünglichen oder nicht mutierten Sequenz (Referenzsequenz / Vergleichsequenz) festgestellt werden können. Sequenzen unter (1) und (2) sind bis auf eine Stelle identisch (unterstrichen). A' und C' sind markiert.



## 1) zu prüfende Sequenz:

Die zu prüfende Sequenz wird mit 2NT's/2NTs-Methode sequenziert, so dass eine Population an NSKF-Sequenzen (ermittelte NSKF-Sequenzen(n)) entsteht. Diese ermittelten NSKF-Sequenzen enthalten Information von jedem Strang:

```
5'A-C---C-AAA-A-C-A-C-CC3' - ermittelte NSKF-Sequenz(i)
5'ATCGTTCGAAATATCGATCGCCTG3'
3'TAGCAAGCTTTATAGCTAGCGGAC5'
3' A-CAA-C---A-A-C-A-C---C5' - ermittelte NSKF-Sequenz(i+1)
```

## 2) Vergleichsequenz:

Zur Analyse ist eine Vergleichsequenz (Referenzsequenz) erforderlich:

```
5'ATTGTTTCGAAATATCGATCGCCTG3'
3'TAACAAGCTTTATAGCTAGCGGAC5'
```

## 3) Vergleichsequenz mit angepaßten ermittelten NSKF-Sequenzen:

Mit Hilfe eines Programms werden ermittelte NSKF-Sequenzen bestimmten Stellen in der Vergleichsequenz zugeordnet und eventuelle Abweichungen detektiert:

```
5'A-C---C-AAA-A-C-A-C-CC3' - ermittelte NSKF-Sequenz(i)
5'ATTGTTTCGAAATATCGATCGCCTG3'
3'TAACAAGCTTTATAGCTAGCGGAC5'
3' A-CAA-C---A-A-C-A-C---C5' - ermittelte NSKF-Sequenz(i+1)
```

↑

(Einzelnukleotidmutation)

Mit dieser Ausführungsform kann man eine doppelsträngige Nukleinsäure auf SNP oder Mutationen untersuchen. Dabei werden die ermittelten NSKF-Sequenzen mit einer Referenzsequenz

verglichen. Die Grundregeln des Vergleichs einer Teilsequenz und einer kompletten Sequenz bei der Analyse mit nur 2 markierten NTs unterscheiden sich nicht prinzipiell von denen, die bei dem Vergleich der Sequenzen anhand aller 4 markierten NTs\* gelten. Näheres s. Sequenzvergleich bei Mutationsanalyse und SNP-Analyse mit 4NTs\* (Beispiel 1B).

### Beispiel 3: Analyse der Genexpression

Die Grundprinzipien der Sequenzierungsreaktion bei der Genexpressionsanalyse entsprechen denen der Sequenzierungsreaktion langer NSKs (Fig. 7). Die Grundprinzipien zur Durchführung eines Reaktionszyklus (die Wahl der NT\*-Struktur, der Polymerase, der Reaktionsbedingungen für die NT\*-Einbaureaktion und die Abspaltungsreaktion), sowie zur Detektion der Signale von eingebauten NT\*, entsprechen denen im Verfahren zur Sequenzierung langer NSKs. Die wesentlichen Unterschiede zwischen beiden Verfahren liegen in der Materialauswahl und -Vorbereitung und in der Verarbeitung der gewonnenen Daten.

#### Auswahl des Materials:

Genprodukte können von verschiedenen biologischen Objekten stammen, so z.B. von einzelnen Zellen, Zellpopulationen, einem Gewebe oder von kompletten Organismen. Auch biologische Flüssigkeiten wie Blut, Sputum oder Liquor können als Quelle der Genprodukte dienen. Die Methoden zur Gewinnung der Genprodukte aus den verschiedenen biologischen Objekten sind beispielsweise folgenden Literaturquellen zu entnehmen: "Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc..

Es kann sowohl die Gesamtheit der isolierten Genprodukte als auch ein durch eine Vorselektion ausgewählter Teil davon in die Sequenzierungsreaktion eingesetzt werden. Durch Vorselektion kann man die Menge der zu analysierenden

Genprodukte reduzieren. Die Vorselektion kann beispielsweise durch molekularbiologische Verfahren wie z.B. PCR-Amplifikation, Gel-Auftrennung oder Hybridisierung mit anderen Nukleinsäureketten erfolgen ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.)

Vorzugsweise wird die Gesamtheit der Genprodukte als Ausgangsmaterial gewählt.

#### Vorbereitung des Materials:

Ziel der Vorbereitung des Materials ist es, aus dem Ausgangsmaterial an die Oberfläche gebundene, extensionsfähige Genprodukt-Primer-Komplexe zu bilden. Wobei pro Genprodukt maximal nur ein Primer binden sollte.

#### Primerbindungsstelle (PBS):

Jedes Genprodukt hat vorzugsweise nur eine Primerbindungsstelle.

Eine Primerbindungsstelle ist ein Sequenzabschnitt, der eine selektive Bindung des Primers an das Genprodukt ermöglichen soll.

Als Primerbindungsstellen können Abschnitte in der Nukleinsäuresequenz dienen, die in den zu analysierenden Sequenzen natürlicherweise vorkommen (z.B. polyA-Strecken in mRNA). Eine Primerbindungsstelle kann auch zusätzlich in das Genprodukt eingeführt werden ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.).

Aus Gründen der Vereinfachung der Analyse kann es wichtig sein, dass eine möglichst einheitliche Primerbindungsstelle in allen Genprodukten vorhanden ist. Dann können Primer mit einheitlicher Struktur in die Reaktion eingesetzt werden.

Die Zusammensetzung der Primerbindungsstelle ist nicht einge-

schränkt. Ihre Länge beträgt vorzugsweise zwischen 10 und 100 NTs. Die Primerbindungsstelle kann eine funktionelle Gruppe tragen, beispielsweise zur Bindung des Genprodukts an die Oberfläche. Diese funktionelle Gruppe kann z.B. eine Biotin- oder Digoxigenin-Gruppe sein.

Als Beispiel für die Einführung einer Primerbindungsstelle in die Genprodukte wird das Nukleotid-Tailing von antisense cDNA-Fragmenten beschrieben.

Als erstes werden einzelsträngige cDNAs von mRNAs synthetisiert. Es resultiert eine Population an cDNA-Molekülen, die eine Kopie der mRNA-Population darstellen, sogenannte antisense-cDNA. (Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory, "Method in Enzymology" 1999, v303, "cDNA library protocols" 1997, Ed. I.G. Cowell, Humana Press Inc.). Mit einer terminalen Deoxynucleotidyltransferase kann man mehrere (z.B. zwischen 10 und 20) Nukleosid-monophosphate an das 3'-Ende dieser antisense cDNA anknüpfen, z.B. mehrere Adenosin-Monophosphate ((dA)n-Tail genannt). Das entstehende Fragment wird zur Bindung des Primers, in diesem Beispiel eines (dT)n-Primers, verwendet ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press, "Method in Enzymology" 1999 v.303, S.37-38).

#### Primer für die Sequenzierungsreaktion:

Dieser hat die Funktion, den Start an einer einzigen Stelle des Genprodukts zu ermöglichen. Vorzugsweise bindet er an die Primerbindungsstelle im Genprodukt. Die Zusammensetzung und die Länge des Primers sind nicht eingeschränkt. Außer der Startfunktion kann der Primer auch andere Funktionen übernehmen, wie z.B. eine Verbindung der Genprodukt-Primer-Komplexe zur Reaktionsoberfläche zu schaffen. Primer sollten so an die Länge und Zusammensetzung der Primerbindungsstelle angepaßt werden, dass der Primer den Start der Sequenzierungsreaktion mit der jeweiligen Polymerase ermöglicht.

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100 NTs, optimalerweise zwischen 15 und 30 NTs. Der Primer kann eine funktionelle Gruppe tragen, die beispielsweise zur Bindung des Primers an die Oberfläche dient, beispielsweise ist eine solche funktionelle Gruppe eine Biotingruppe (s. Abschnitt Immobilisierung). Sie soll die Sequenzierung nicht stören. Die Synthese eines solchen Primers kann z.B. mit dem DNA-Synthesizer 380 A Applied Biosystems ausgeführt werden oder aber als Auftragssynthese bei einem kommerziellen Anbieter, z.B. MWG-Biotech GmbH, Deutschland, erstellt werden.

Es können auch unterschiedliche Primer verwendet werden, ein definierter Primersatz, oder ein Primergemisch.

Der Primer kann vor der Hybridisierung an die zu analysierenden Fragmente auf der Oberfläche mit verschiedenen Techniken fixiert oder direkt auf der Oberfläche synthetisiert werden beispielsweise nach (McGall et al. US Patent 5412087, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Fodor et al. Science 1991 v.285 S.767, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679).

Die Primer werden auf der Oberfläche in einer Dichte zwischen 10 bis 100 pro 100  $\mu\text{m}^2$ , 100 bis 10.000 pro 100  $\mu\text{m}^2$ , 10.000 bis 1.000.000 pro 100  $\mu\text{m}^2$  oder größer als 1.000.000 pro 100  $\mu\text{m}^2$  gebunden.

Der Primer oder das Primergemisch wird mit Genprodukten unter Hybridisierungsbedingungen inkubiert, die ihn selektiv an die Primerbindungsstelle jedes Genprodukts binden lassen. Diese Primer-Hybridisierung (Annealing) kann vor (1), während (2) oder nach (3) der Bindung der Genprodukte an die Oberfläche

erfolgen. Falls Genprodukte als doppelsträngige Nukleinsäuren vorliegen, werden sie vor der Hybridisierung durch Hitze denaturiert ("Molecular cloning" 1989 J.Sambrook et al. Cold Spring Harbor Laboratory Press). Die Optimierung der Hybridisierungsbedingungen hängt von der genauen Struktur der Primerbindungsstelle und des Primers ab und läßt sich nach Rychlik et al. (NAR 1990 v.18 S.6409) berechnen. Im folgenden werden diese Hybridisierungsbedingungen als standardisierte Hybridisierungsbedingungen bezeichnet.

Falls das Ausgangsmaterial eine poly-A-Strecke oder eine poly-dA-Strecke aufweist (z.B. mRNA, sense cDNA oder antisense-cDNA mit (dA)<sub>n</sub>-Tail) kann man einen oligo-dT-Primer verwenden. Es kann allerdings auch ein Primergemisch bestehend aus 12 verschiedenen Primern mit folgender allgemeiner Struktur 5'(K)<sub>n</sub>MN3' verwendet werden. Wobei (n) zwischen 10 und 50 liegt, vorzugsweise zwischen 20 und 30. "K" steht für dT oder dU, "M" und "N" stehen jeweils für dA, dT oder dU, dC, dG (z.B. 5'-dTdTdTdTdTdTdTdTdT<sub>10</sub>dTdTdTdTdTdTdTdTdT<sub>20</sub>dAdG-3'). Ein solches Primergemisch ermöglicht einen exakten Start der Sequenzierungsreaktion am Ende der polyA-Strecke oder der poly-dA-Strecke (geankerter Primer).

Fixierung von Genprodukt-Primer-Komplexen an die Oberfläche (Bindung bzw. Immobilisierung von Genprodukten):

Ziel der Fixierung (Bindung, Immobilisierung) ist es, Genprodukt-Primer-Komplexe auf einer geeigneten planen Oberfläche in einer Art und Weise zu fixieren, dass eine zyklische enzymatische Sequenzierungsreaktion ablaufen kann. Dies kann beispielsweise durch Bindung des Primers (s.o.) oder des Genprodukts an die Oberfläche erfolgen.

Die Reihenfolge der Schritte bei der Bindung von Genprodukt-Primer-Komplexen kann variabel sein:

- 4) Die Genprodukt-Primer-Komplexe können zunächst in einer Lösung durch Hybridisierung (Annealing) gebildet und

anschließend an die Oberfläche gebunden werden.

- 5) Primer können zunächst auf einer Oberfläche gebunden werden und Genprodukte anschließend an die gebundenen Primer hybridisiert werden, wobei Genprodukt-Primer-Komplexe entstehen (Genprodukte indirekt an die Oberfläche gebunden)
- 6) Die Genprodukte können zunächst an die Oberfläche gebunden werden (Genprodukte direkt an die Oberfläche gebunden) und im anschließenden Schritt die Primer an die gebundenen Genprodukte hybridisiert werden, wobei Genprodukt-Primer-Komplexe entstehen.

Die Immobilisierung der Genprodukte an die Oberfläche kann daher durch direkte oder indirekte Bindung erfolgen.

Oberfläche und Reaktionsoberfläche sind in dieser Anmeldung als gleichwertige Begriffe aufzufassen, außer wenn explizit auf eine andere Bedeutung hingewiesen wird. Als Reaktionsoberfläche dient die Oberfläche einer festen Phase eines beliebigen Materials. Dieses Material ist vorzugsweise enzymatischen Reaktionen gegenüber inert und verursacht keine Störungen der Detektion. Silicon, Glas, Keramik, Kunststoff (z.B. Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material, das diesen funktionalen Anforderungen genügt, kann verwendet werden. Vorzugsweise ist die Oberfläche nicht verformbar, denn sonst ist mit einer Verzerrung der Signale bei der wiederholten Detektion zu rechnen.

Falls eine gelartige feste Phase (Oberfläche eines Gels) verwendet wird, so kann dieses Gel z.B. ein Agarose- oder Polyacrylamidgel sein. Das Gel ist vorzugsweise für Moleküle mit einer Molekularmasse unter 5000 Da frei passierbar (beispielsweise kann ein 1 bis 2% Agarose-Gel oder 5 bis 15% Polyacrylamid Gel verwendet werden). Eine solche Geloberfläche hat anderen festen Oberflächen gegenüber den Vorteil, dass es zu einer wesentlich geringeren unspezifischen Bindung von NT\*s an die Oberfläche kommt. Durch die Bindung der Genprodukt-

Primer-Komplexe auf der Oberfläche ist die Detektion der Fluoreszenzsignale von eingebauten NTs\* möglich. Die Signale von freien NTs\* werden nicht detektiert, weil sie nicht an das Material des Gels binden und somit nicht immobilisiert werden. Das Gel ist vorzugsweise auf einer festen Unterlage befestigt. Diese feste Unterlage kann Silicon, Glas, Keramik, Kunststoff (z.B. Polycarbonate oder Polystyrole), Metall (Gold, Silber, oder Aluminium) oder beliebiges anderes Material sein.

Die Dicke des Gels beträgt vorzugsweise nicht mehr als 0,1 mm. Die Geldicke ist jedoch vorzugsweise größer als die einfache Tiefenschärfe des Objektivs, damit unspezifisch an die feste Unterlage gebundene NTs\* nicht in die Fokusebene gelangen und damit detektiert werden. Wenn die Tiefenschärfe z.B. 0,3  $\mu\text{m}$  beträgt, so liegt die Geldicke vorzugsweise zwischen 1  $\mu\text{m}$  und 100  $\mu\text{m}$ . Die Oberfläche kann als eine kontinuierliche Oberfläche oder als diskontinuierliche, aus einzelnen kleinen Bestandteilen (z.B. Agarose-Kügelchen) zusammengesetzte Oberfläche hergestellt werden. Die Reaktionsoberfläche muß groß genug sein, um die notwendige Anzahl der Genprodukte bei entsprechender Dichte binden zu können. Die Reaktionsoberfläche sollte vorzugsweise nicht größer als 20  $\text{cm}^2$  sein.

Die verschiedenen Zyklusschritte erfordern einen Austausch der unterschiedlichen Reaktionslösungen über der Oberfläche. Die Reaktionsoberfläche ist vorzugsweise Bestandteil eines Reaktionsgefäßes. Das Reaktionsgefäß ist wiederum vorzugsweise Bestandteil einer Reaktionsapparatur mit Durchflußvorrichtung. Die Durchflußvorrichtung ermöglicht einen Austausch der Lösungen im Reaktionsgefäß. Der Austausch kann mit einer durch einen Computer gesteuerten Pumpvorrichtung oder manuell erfolgen. Wichtig dabei ist, dass die Oberfläche nicht austrocknet. Vorzugsweise beträgt das Volumen des Reaktionsgefäßes weniger als 50  $\mu\text{l}$ . Idealerweise beträgt sein Volumen weniger als 1  $\mu\text{l}$ .

Falls die Fixierung der Genprodukt-Primer-Komplexe auf der Oberfläche über die Genprodukte erfolgt, kann dies



beispielsweise durch die Bindung der Genprodukte an einem der beiden Ketten-Enden erfolgen. Dies kann durch entsprechende kovalente, affine oder andere Bindungen erreicht werden. Es sind viele Beispiele der Immobilisierung von Nukleinsäuren bekannt (McGall et al. US Patent 5412087, Nikiforov et al. US Patent 5610287, Barrett et al. US Patent 5482867, Mirzabekov et al. US Patent 5981734, "Microarray biochip technology" 2000 M.Schena Eaton Publishing, "DNA Microarrays" 1999 M. Schena Oxford University Press, Rasmussen et al. Analytical Biochemistry v.198, S.138, Allemand et al. Biophysical Journal 1997, v.73, S.2064, Trabesinger et al. Analytical Chemistry 1999, v.71, S.279, Osborne et al. Analytical Chemistry 2000, v.72, S.3678, Timofeev et al. Nucleic Acid Research (NAR) 1996, v.24 S.3142, Ghosh et al. NAR 1987 v.15 S.5353, Gingeras et al. NAR 1987 v.15 S.5373, Maskos et al. NAR 1992 v.20 S.1679). Die Fixierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der Genprodukte enthaltenden Probe auf der planen Oberfläche erreicht werden. Die Genprodukte werden auf der Oberfläche in einer Dichte zwischen 10 und 100 pro 100  $\mu\text{m}^2$ , 100 bis 10.000 pro 100  $\mu\text{m}^2$ , 10.000 bis 1000.000 pro 100  $\mu\text{m}^2$  gebunden.

Die für die Detektion notwendige Dichte von extensionsfähigen Genprodukt-Primer-Komplexen beträgt ca. 10 bis 100 pro 100  $\mu\text{m}^2$ . Sie kann vor, während oder nach der Hybridisierung der Primer an die Genprodukte erreicht werden.

Beispielhaft werden im folgenden einige Methoden zur Bindung näher dargestellt:

In einer Ausführungsform erfolgt die Bindung der Genprodukt-Primer-Komplexe über Biotin-Avidin oder Biotin-Streptavidin-Bindung. Dabei wird Avidin oder Streptavidin auf der Oberfläche kovalent gebunden, das 5'-Ende des Primers ist mit Biotin modifiziert. Nach der Hybridisierung der modifizierten Primer mit den Genprodukten (in Lösung) werden diese auf der mit Avidin/Streptavidin beschichteten Oberfläche fixiert. Die Konzentration der mit Biotin markierten Genprodukt-Primer-

Komplexe sowie die Zeit der Inkubation dieser Lösung mit der Oberfläche wird so gewählt, dass eine für die Sequenzierung geeignete Dichte erreicht wird.

In einer anderen Ausführungsform werden die für die Sequenzierungsreaktion geeigneten Primer vor der Sequenzierungsreaktion auf der Oberfläche mit geeigneten Methoden fixiert (s.o.). Die einzelsträngigen Genprodukte mit jeweils einer Primerbindungsstelle pro Genproduktmolekül werden damit unter Hybridisierungsbedingungen inkubiert (Annealing). Dabei binden sie an die fixierten Primer und werden dadurch an die Oberfläche gebunden (indirekte Bindung). Die Konzentration der einzelsträngigen Genprodukte und die Hybridisierungsparameter (z.B. Temperatur, Zeit, Puffer) werden so gewählt, dass man eine für die Sequenzierung geeignete Dichte von ca. 10 bis 100 extensionsfähigen Genprodukt-Primer-Komplexen pro 100  $\mu\text{m}^2$  erreicht. Nach der Hybridisierung werden ungebundene Genprodukte durch einen Waschschrift entfernt. Bei dieser Ausführungsform wird eine Oberfläche mit einer hohen Primerdichte bevorzugt, z.B. ca. 1.000.000 Primer pro 100  $\mu\text{m}^2$  oder noch höher, da die gewünschte Dichte an Genprodukt-Primer-Komplexen schneller erreicht wird, wobei die Genprodukte nur an einen Teil der Primer binden.

In einer anderen Ausführungsform werden die Genprodukte an die Oberfläche direkt gebunden (s.o.) und anschließend mit Primern unter Hybridisierungsbedingungen inkubiert. Bei einer Dichte von ca. 10 bis 100 Genprodukte pro 100  $\mu\text{m}^2$  wird man versuchen alle verfügbaren Genprodukte mit einem Primer zu versehen und für die Sequenzierungsreaktion verfügbar zu machen. Dies kann durch hohe Primerkonzentration beispielsweise 1 bis 100 mmol/l erreicht werden. Bei einer höheren Dichte der fixierten Genprodukte auf der Oberfläche, beispielsweise 10.000 bis 1.000.000 pro 100  $\mu\text{m}^2$ , kann die für die optische Detektion notwendige Dichte der Genprodukt-Primer-Komplexe während der Primer-Hybridisierung erreicht werden. Dabei sind die Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer,

Primerkonzentration) so zu wählen, dass die Primer nur an einen Teil der immobilisierten Genprodukte binden.

Falls die Oberfläche einer festen Phase (z.B. Silikon oder Glas) zur Immobilisation verwendet wird, wird vorzugsweise eine Blockierungslösung auf die Oberfläche vor dem Schritt (a) in jedem Zyklus gebracht, die zur Vermeidung einer unspezifischen Adsorption von NTs\* an der Oberfläche dient. Diese Bedingungen für eine Blockierungslösung erfüllt beispielsweise eine Albuminlösung (BSA) mit einem pH-Wert zwischen 8 und 10.

#### Wahl der Polymerase:

Bei der Wahl der Polymerase spielt die Art der verwendeten immobilisierten Nukleinsäure (RNA oder DNA) eine entscheidende Rolle:

Falls RNA als Genprodukt (z.B. mRNA) in die Sequenzierungsreaktion eingesetzt wird, können handelsübliche RNA-abhängige DNA-Polymerasen eingesetzt werden, z.B. AMV-Reverse Transcriptase (Sigma), M-MLV Reverse Transcriptase (Sigma), HIV-Reverse Transcriptase ohne RNase-Aktivität. Alle Reverse Transcriptasen müssen von RNase-Aktivität weitgehend frei sein ("Molecular cloning" 1989, Ed. Maniatis, Cold Spring Harbor Laboratory).

Falls DNA als Genprodukt (z.B. cDNA) verwendet wird, eignen sich als Polymerasen prinzipiell alle DNA-abhängigen DNA-Polymerasen ohne 3'-5' Exonuklease-Aktivität (DNA-Replication" 1992 Ed. A.Kornberg, Freeman and company NY), z.B. modifizierte T7-Polymerase vom Typ "Sequenase Version 2" (Amersham Pharmacia Biotech), Klenow Fragment der DNA-Polymerase I ohne 3'-5' Exonukleaseaktivität (Amersham Pharmacia Biotech), Polymerase Beta verschiedenen Ursprungs (Animal Cell DNA Polymerases" 1983, Fry M., CRC Press Inc., kommerziell erhältlich bei Chimerx) thermostabile Polymerasen wie Taq-Polymerase (GibcoBRL), proHA-DNA-Polymerase (Eurogentec).

Detektion:

Wie bei der Sequenzierung langer NSKs, umfasst die Detektion folgende Phasen:

## 1) Vorbereitung zur Detektion

Durchführung eines Detektionsschrittes in jedem Zyklus, wobei jeder Detektionsschritt als Scanvorgang abläuft und folgende Operationen umfaßt:

- a) Einstellung der Position des Objektivs (X,Y-Achse),
- b) Einstellung der Fokusebene (Z-Achse),
- c) Detektion der Signale einzelner Moleküle, Zuordnung des Signals zu NT\* und Zuordnung des Signals zum jeweiligen Genprodukt,
- d) Verschiebung zur nächsten Position auf der Oberfläche.

Die Signale von in die den Genprodukten komplementären Stränge eingebauten NTs\* werden durch das Abscannen der Oberfläche registriert. Der Scanvorgang wird wie bei der Sequenzierung langer NSKs durchgeführt. Dabei wird das Objektiv schrittweise über die Oberfläche bewegt, so daß von jeder Oberflächenposition ein zweidimensionales Bild (2D-Bild) entsteht.

Vorbereitung zur Detektion:

Am Anfang wird festgelegt, wie viele Kopien der Genprodukte zur Expressionsanalyse notwendig sind. Mehrere Faktoren spielen dabei eine Rolle. Die genaue Zahl hängt z.B. von der relativen Präsenz der Genprodukte im Ansatz und von der gewünschten Genauigkeit der Analyse ab. Die Anzahl der analysierten Genprodukte liegt vorzugsweise zwischen 1000 und 10.000.000. Für stark exprimierte Gene kann die Anzahl der analysierten Genprodukte niedrig sein, z.B. 1000 bis 10.000. Bei der Analyse schwach exprimierter Gene muß sie erhöht werden, z.B. auf 100.000 oder noch weiter.

Es werden beispielsweise 100.000 einzelne Genprodukte gleichzeitig analysiert. Dabei werden auch schwach exprimierte Gene (mit z.B. ca.100 mRNA-Molekülen/Zelle, was

ca. 0.02% gesamt-mRNA entspricht) in der Reaktion mit durchschnittlich 20 identifizierten Genprodukten repräsentiert.

2) Durchführung eines Detektionsschrittes in jedem Zyklus gleicht der in der Sequenzierung langer NSKs. Anstelle von NSKFs werden Genprodukte verwendet.

#### Analyse:

Die gewonnenen Daten (kurze Sequenzen) werden mit Hilfe eines Programms mit bekannten Gensequenzen verglichen. Einem solchen Programm kann z.B. ein BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Durch die Wahl der Methode zur Materialvorbereitung wird unter anderem festgelegt, in welchen Abschnitten der Genprodukte die Sequenzen ermittelt werden und zu welchem Strang (sense oder antisense) sie gehören. Z.B. werden bei der Verwendung der polyA-Strecken als Primerbindungsstelle in mRNA Sequenzen aus NTRs (non-translating-regions) bestimmt. Bei der Verwendung der Methode mit antisense-cDNA als Matrize stammen die ermittelten Sequenzen unter anderem aus den proteinkodierenden Bereichen der Genprodukte.

Bei einer bevorzugten einfachen Variante der Erfindung wird die Genexpression nur qualitativ bestimmt. Dabei ist nur die Tatsache der Expression bestimmter Gene von Bedeutung.

Bei einer anderen bevorzugten Ausführungsform ist eine quantitative Bestimmung der Verhältnisse zwischen einzelnen Genprodukten im Ansatz von Interesse. Es ist bekannt, daß die Aktivität eines Gens in einer Zelle durch eine Population identischer mRNA-Moleküle repräsentiert ist. In einer Zelle sind viele Gene gleichzeitig aktiv und werden dabei unterschiedlich stark exprimiert, was zum Vorhandensein vieler verschiedener unterschiedlich stark repräsentierter mRNA-Populationen führt.

Im folgenden wird auf die quantitative Analyse der Genexpression näher eingegangen:

Für eine quantitative Analyse der Genexpression werden die Abundanzen einzelner Genprodukte in der Sequenzierungsreaktion bestimmt. Dabei sind die Produkte stark exprimierter Gene in der Sequenzierungsreaktion häufiger vertreten als die schwach exprimierter Gene.

Nach der Zuordnung der Sequenzen zu bestimmten Genen wird der Anteil der ermittelten Sequenzen für jedes einzelne Gen bestimmt. Gene mit starker Expression haben einen höheren Anteil an der Gesamtpopulation der Genprodukte als Gene mit schwacher Expression.

Die Anzahl der analysierten Genprodukte liegt vorzugsweise zwischen 1000 und 10.000.000. Die genaue Anzahl der zu analysierenden Genprodukte hängt von der Aufgabenstellung ab. Für stark exprimierte Gene kann sie niedrig sein, z.B. 1000 bis 10.000. Bei der Analyse schwach exprimierter Gene muß sie erhöht werden, z.B. auf 100.000 oder höher.

Werden beispielsweise 100.000 einzelne Genprodukte gleichzeitig analysiert, sind auch schwach exprimierte Gene, wie z.B. ca. 100 mRNA-Moleküle/Zelle (was ca. 0.02% gesamt-mRNA entspricht), in der Reaktion mit durchschnittlich 20 identifizierten Genprodukten repräsentiert.

Als interne Kontrolle der Hybridisierung, der Immobilisation und der Sequenzierungsreaktion läßt sich folgende Methode verwenden:

Es können eine oder mehrere Nukleinsäureketten mit bekannten Sequenzen als Kontrolle eingesetzt werden. Die Zusammensetzung dieser Kontrollsequenzen ist nicht eingeschränkt, sofern sie die Identifizierung der Genprodukte nicht stören. Bei der Sequenzanalyse der mRNA-Proben werden RNA-Kontrollproben, bei der Analyse der cDNA-Proben entsprechend DNA-Kontrollproben eingesetzt. Diese Proben werden vorzugsweise bei allen Schritten mitgeführt. Sie

können z.B. nach der mRNA-Isolation zugegeben werden. Im allgemeinen werden die Kontrollproben in gleicher Weise zur Sequenzanalyse vorbereitet wie die zu analysierenden Genprodukte.

Die Kontrollsequenzen werden in bekannten, fest eingestellten Konzentrationen zu den zu analysierenden Genprodukten zugegeben. Konzentrationen der Kontrollproben können unterschiedlich sein, vorzugsweise liegen diese Konzentrationen zwischen 0.01% und 10% der Gesamtkonzentration der zu analysierenden Probe (100%). Beträgt die Konzentration der mRNA beispielsweise 10ng/ $\mu$ l, dann liegen die Konzentrationen von Kontrollproben zwischen 1pg/ $\mu$ l und 1ng/ $\mu$ l.

Bei der quantitativen Analyse der Genexpression muß auch die allgemeine metabolische Aktivität der Zellen berücksichtigt werden, insbesondere, wenn ein Vergleich der Expression bestimmter Gene bei verschiedenen äußeren Bedingungen angestrebt wird.

Die Veränderung im Expressionsniveau eines bestimmten Gens kann als Folge der Veränderung in der Transkriptionsrate dieses Gens oder als Folge einer globalen Veränderung der Genexpression in der Zelle auftreten. Zur Beobachtung der metabolischen Zustände in der Zelle kann man die Expression der sogenannten "House-keeping-Gene" analysieren. Beim Mangel an wichtigen Metaboliten ist beispielsweise das allgemeine Expressionsniveau in der Zelle niedrig, so daß auch konstitutiv exprimierte Gene ein niedriges Expressionsniveau haben.

Im Prinzip können alle konstitutiv exprimierten Gene als "House-keeping-Gene" dienen. Als Beispiele seien das Transferrin-Rezeptor-Gen oder das Beta-Aktin-Gen genannt.

Die Expression dieser House-keeping-Gene dient somit als Bezugsgröße für die Analyse der Expression anderer Gene. Die Sequenzermittlung und Quantifizierung der Expression der House-keeping-Gene ist vorzugsweise ein Bestandteil des  
5 Analyse-Programms für die Genexpression.

Wie bei der Sequenzierung langer NSKs kann man Sequenzierungsreaktion mit 4 markierten oder 2 markierten und 2 unmarkierten NT\* durchführen.

10

Sequenzanalyse mit 4 markierten NTs\*Sequenzanalyse mit 4 markierten NTs\*

Bei einer bevorzugten Ausführungsform der Erfindung werden  
15 alle vier in die Reaktion eingesetzten NTs\* mit Fluoreszenzfarbstoffen markiert.

Dabei verwendet man eine der oben genannten farbigen Kodierungsschemata. Die Zahl der ermittelten NTs für jede Sequenz aus einem Genprodukt liegt zwischen 5 und 100,  
20 idealerweise zwischen 20 und 50. Diese ermittelten Sequenzen werden mit Hilfe eines Programms mit bekannten Sequenzen in Gen-Datenbanken verglichen und entsprechenden Genen zugeordnet. Einem solchen Programm kann z.B. der BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to  
25 computational Biology" 1995 M.S. Waterman Chapman & Hall).

Ein Zyklus hat folgende Schritte:

- a) Zugabe einer Lösung mit markierten Nukleotiden (NTs\*)  
und Polymerase zu immobilisierten  
30 Nukleinsäureketten,
- b) Inkubation der immobilisierten Nukleinsäureketten mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- 35 c) Waschen
- d) Detektion der Signale von einzelnen Molekülen
- e) Entfernung der Markierung von den eingebauten Nukleotiden und des zur Termination führenden



Substituenten,  
f) Waschen.

Sequenzanalyse mit 2 markierten NTs\* und 2 unmarkierten NTs  
(2NTs\* / 2NTs-Methode).

In einer anderen Ausführungsform werden für die Analyse der Sequenzen 2 modifizierte NTs\* und 2 unmodifizierte NTs eingesetzt.

Diese Ausführungsform beruht auf dem Prinzip, daß eine Abfolge aus 2 Signalen (markierte NT\*s) genügend Informationen zur Identifizierung einer Sequenz enthalten kann. Die ermittelte Sequenz wird mit der Referenzsequenz verglichen und einer bestimmten Position zugeordnet, z.B.:

ACCAAAACACCC - ermittelte Sequenz (dCTP\* und dATP\* sind markiert)

ATCATCGTTTCGAAATATCGATCGCCTGATGCC - Referenzsequenz

A-C---C-AAA-A-C-A-C-CC (zugeordnete ermittelte Sequenz)  
ATCATCGTTTCGAAATATCGATCGCCTGATGCC (Referenzsequenz)

Vorzugsweise wird die ermittelten Sequenzen mit Hilfe eines Programms der Referenzsequenz zugeordnet. Einem solchen Programm kann z.B. der BLAST oder FASTA Algorithmus zugrunde liegen ("Introduction to computational Biology" 1995 M.S. Waterman Chapman & Hall).

Die Genprodukte werden wie oben beschrieben zur Sequenzierung vorbereitet und mit 2NTs\*/2NTs-Methode sequenziert. Man erhält Sequenzenabschnitte aus Genprodukten, wobei jede Sequenz eine Abfolge aus 2NTs\* darstellt. Bekannte Gensequenzen dienen als Referenzsequenzen. Um eine eindeutige

Zuordnung der ermittelten Sequenz zu einer bekannten Referenzsequenz zu ermöglichen, muß diese Abfolge lang genug sein. Vorzugsweise beträgt die Länge der ermittelten Sequenzen mehr als 20 NT's. Da 2 markierte NTs' nur einen Teil der Sequenz darstellen, ist die Gesamtlänge des synthetisierten komplementären Strangs ca. doppelt so lang, wie die Abfolge der detektierten NTs' (bei 20 detektierten NTs' beträgt die Gesamtlänge z.B. durchschnittlich 40 NTs).

Zur Synthese eines komplementären Stranges werden 4 Nukleotide benötigt. Da die mit einem Fluoreszenzfarbstoff markierten NTs' in der vorliegenden Erfindung als Semiterminatoren auftreten, d.h. die Termination ausschließlich bei Verfügbarkeit modifizierter NTs' auftritt, müssen unmodifizierte NTs in einem zusätzlichen Schritt in jedem Zyklus in die Reaktion zugegeben werden. Die genaue Position dieses Schrittes in dem Zyklus kann variieren. Wichtig dabei ist, daß die markierten NTs' und die unmodifizierte NTs getrennt verwendet werden.

Ein Zyklus bei dieser Ausführungsform kann beispielhaft folgendermaßen aussehen:

- a) Zugabe einer Lösung mit modifizierten NTs' und Polymerasen auf die Oberfläche mit den bereitgestellten Genprodukten
- b) Inkubation der immobilisierten Nukleinsäureketten mit dieser Lösung unter Bedingungen, die zur Verlängerung der komplementären Stränge um ein NT geeignet sind,
- c) Waschen
- d) Detektion der Signale von einzelnen, modifizierten und in die den Genprodukten komplementären neusynthetisierten Strängen eingebauten NTs'-Molekülen
- e) Entfernung der Markierung und der terminierenden Gruppe bei den eingebauten Nukleotiden
- f) Waschen

- g) Zugabe von 2 unmodifizierten NTs und Polymerasen
- h) Waschen.

Die Konzentration der NTs liegt vorzugsweise unter 1 mM, idealerweise unter 10  $\mu$ M.

#### Beispiel 4:

Eine besondere Ausführungsform des Verfahrens stellt die Analyse von Einzelnukleotidpolymorphismen mit sequenzspezifischen Primern dar.

Zusätzlich zu Abschnitt 1 "Abkürzungen und Begriffserläuterungen" werden für dieses Beispiel folgende Begriffe definiert:

Primer - Zur Verdeutlichung des erfinderischen Gedankens werden in diesem Beispiel folgende Begriffe unterschieden:

- a) Unter einem ~~Primer~~ wird vorliegend allgemein eine Population von Primermolekülen mit einheitlicher Struktur verstanden.
- b) ~~Mehrere Primer~~ o.ä. werden im Text als mehrere Populationen von Primermolekülen verstanden, die unterschiedliche Struktur besitzen.
- c) Ein ~~Primer-Molekül~~ bedeutet ein einziges Oligonukleotid-Molekül.
- d) ~~Mehrere Primer-Moleküle~~ bedeuten mehrere einzelne Oligonukleotid-Moleküle; sie können einheitliche oder unterschiedliche Struktur aufweisen.

SNP-Stelle - eine Position in NSK, die auf Vorhandensein oder Abwesenheit von SNP untersucht wird.

Zielsequenz - Teil einer Gesamtsequenz, der durch die Verwendung eines spezifischen Primers in der Sequenzierungsreaktion sequenziert/ bestimmt wird. Eine

Gesamtsequenz kann mehrere Zielsequenzen enthalten. Eine Zielsequenz ist genügend lang, um eine Positionierung dieser Zielsequenz innerhalb der Gesamtsequenz mit großer Wahrscheinlichkeit zu gewährleisten. Zielsequenzen können beispielsweise eine oder mehrere SNP-Stellen enthalten.

Erkennungssequenz - Teil der Zielsequenz, der für die Zuordnung dieser Zielsequenz in der Gesamtsequenz verwendet wird.

In dieser Ausführungsform zur SNP-Analyse werden mehrere potentielle SNP-Positionen in der Referenzsequenz ausgewählt, die in einer zu analysierenden NSK untersucht werden. Zu diesen Positionen werden entsprechend unterschiedliche, sequenzspezifische Primer bereitgestellt. Diese Primer können einen standardisierten Primersatz zur SNP-Analyse bei einer bestimmten Fragestellung bilden und einheitlich als Kit für die betreffende Analysen eingesetzt werden.

Die Vorbereitung des zu analysierenden Materials (auf SNP zu untersuchende einzel- und doppelsträngige Nukleinsäureketten) hat erfindungsgemäß das Ziel, aus einer oder mehreren langen Nukleinsäureketten (Gesamtsequenz) eine Population an relativ kleinen, zwischen 30 und 2000 NT langen, einzelsträngigen Nukleinsäurekettenfragmenten (NSKFs) zu bilden.

Diese NSKF-Moleküle werden zufällig auf einer planen Oberfläche mit einer Dichte zwischen 10 und 1.000.000 pro 100  $\mu\text{m}^2$ , vorzugsweise 10 und 100 NSKFs pro 100  $\mu\text{m}^2$ , 100 bis 10.000 pro 100  $\mu\text{m}^2$  oder 10.000 bis 1.000.000 pro 100  $\mu\text{m}^2$  immobilisiert. An die auf der Oberfläche gebundenen NSKFs werden Primer hybridisiert, so dass die Dichte der extensionsfähigen NSKF-Primer-Komplexe ca. 10-100 pro 100  $\mu\text{m}^2$  beträgt. Nach der Hybridisierung werden nicht gebundene Primer entfernt und die Sequenzierungsreaktion gestartet.

Durch eine Auswahl der Zielsequenzen und der

sequenzspezifischen Primer werden nur die relevanten Abschnitte der Gesamtsequenz untersucht, was die Menge nicht relevanter Informationen verringert und die Analysezeit verkürzt.

Dieser Ausführungsform des Verfahrens zur SNP-Analyse liegen folgende Prinzipien zugrunde:

Es werden Stellen in einer Referenzsequenz ausgewählt, die in den zu untersuchenden NSKs (Gesamtsequenz) auf Einzelnukleotidpolymorphismen (SNPs) überprüft werden sollen.

- 1) Zur Analyse jeder ausgewählten SNP-Stelle werden spezifische Primer bereitgestellt, so dass jede zu untersuchende SNP-Stelle entweder die nächste Position in 3'-Richtung vom Primer einnimmt oder innerhalb von 2 bis 100, vorzugsweise 2 bis 50, idealerweise 2 bis 20 Positionen in 3'-Richtung vom Primer liegt. Die SNP-Stelle liegt somit innerhalb der Zielsequenz, die während der Sequenzierungsreaktion bestimmt wird. Es werden vorzugsweise mehrere SNP-Stellen gleichzeitig analysiert, so dass mehrere spezifische Primer verwendet werden müssen. Die Primer werden vorzugsweise so ausgewählt, dass sie möglichst einheitliche Annealing-Temperaturen haben, d.h. Unterschiede zwischen Schmelztemperaturen einzelner Primerpopulationen liegen beispielsweise innerhalb eines Bereichs von ca. 4 Grad, besser innerhalb von 2 Grad, noch besser innerhalb von 1 Grad.
- 2) Von der Gesamtsequenz werden kurze Nukleinsäurekettenfragmente (NSKFs) abgeleitet, wobei diese Fragmente einzelsträngig sind und eine Länge von 20 bis 2000 NT, vorzugsweise 30 bis 500 NT besitzen.
- 3) NSKF-Moleküle werden in einer zufälligen Anordnung auf

der Oberfläche immobilisiert.

- 4) Nach der Hybridisierung (Annealing) von sequenzspezifischen Primern an die auf der Oberfläche immobilisierten NSKFs wird eine zyklische Sequenzierungsreaktion durchgeführt, wobei für jedes an der Reaktion beteiligte NSKF-Molekül eine Zielsequenz ermittelt wird. Die Sequenzierungsreaktion läuft an vielen Molekülen gleichzeitig ab.
- 5) Die ermittelten Zielsequenzen enthalten Information über die Zugehörigkeit zu einem bestimmten Abschnitt in der Gesamtsequenz und über den SNP in diesem Abschnitt bei der zu untersuchenden Probe. Die Länge der Zielsequenzen und somit die Zahl der Zyklen ist so zu wählen, dass eine Identifizierung der Sequenzen gewährleistet werden kann.

In einer vorteilhaften Ausführungsform werden die ermittelten Zielsequenzen mit der Referenzsequenz verglichen und durch Sequenzübereinstimmung zugeordnet. Bei einer genügend langen ermittelten Zielsequenz kann man sie mit großer Wahrscheinlichkeit zu einer bestimmten Position in der Referenzsequenz zuordnen. Beispielsweise kann eine Sequenz aus 10 NTs mehr als  $10^6$  verschiedene Kombinationen bilden und somit mit einer großen Wahrscheinlichkeit in einer NSK von nur 100.000 NT eindeutig identifiziert werden. Nach der Zuordnung der ermittelten Zielsequenz zur bestimmten Position innerhalb der Referenzsequenz werden Unterschiede in den Sequenzen, die SNPs, sichtbar.

Zur Identifizierung der Zielsequenzen wird in einer anderen vorteilhaften Ausführungsform sowohl die bereits bekannte Anzahl der Primer, ihre Zusammensetzung und ein bereits bekannter, an die Primerbindungsstelle anschließender Sequenzabschnitt der Referenzsequenz

verwendet. Dabei werden die ermittelten Zielsequenzen nach ihrer Zugehörigkeit zu den Primern analysiert, wobei nur die nah an der Primerbindungsstelle liegenden Sequenzen berücksichtigt werden müssen. Wenn beispielsweise nur 1000 Primer verwendet werden, reichen weniger als 10 NTs der ermittelten Zielsequenzen, um eine Zuordnung zu den entsprechenden Primern zu ermöglichen.

Die zu analysierende Probe enthält meistens mehrere identische Gesamtsequenzmoleküle, z.B. mehrere Kopien von genomischer DNA aus Zellen eines Gewebes oder mehrere identische mRNA-Populationen aus Zellen eines Gewebes.

*a) Wahl der SNP-Stelle*

Mit der erfindungsgemäßen Methode kann man sowohl bekannte SNP-Stellen analysieren als auch neue SNP-Stellen ermitteln. Als potentielle SNP-Stelle kann jede Position in der NSK auftreten. Die Auswahl richtet sich nach der Fragestellung, z.B. SNP-Analyse in Genen, deren Produkte mit bestimmten Krankheiten assoziiert sind, oder SNP-Analyse in konservierten, kodierenden Abschnitten der Gene, die für Membranrezeptoren kodieren, oder Überprüfung bekannter SNP-Stellen in regulatorischen Sequenzen von Genen, die für die Zellteilung wichtig sind.

Eine zu analysierende SNP-Stelle liegt innerhalb einer Zielsequenz, die während der Sequenzierungsreaktion bestimmt wird. Man kann mehrere SNP-Stellen innerhalb einer Zielsequenz ermitteln. Man kann andererseits auch mehrere Zielsequenzen z.B. innerhalb eines Gens wählen. Wichtig dabei ist, dass die Zielsequenzen in genügendem Abstand voneinander in der Gesamtsequenz liegen. Dieser Abstand ist notwendig, damit nur ein sequenzspezifischer Primer pro NSKF hybridisiert, und er ist von der durchschnittlichen NSKF-Länge abhängig: je kürzer die NSKFs, desto näher aneinander können Zielsequenzen liegen. Die SNP-Stellen können bei angemessener Primer-Wahl an beiden

Strängen einer doppelsträngigen Nukleinsäurekette analysiert werden.

Das Verfahren bietet auch die Möglichkeit, beispielsweise mehrere SNP-Stellen aus vielen Individuen (als Stichprobe einer Population) gleichzeitig zu kontrollieren. Dadurch kann z.B. das SNP-Profil einer Population untersucht werden.

*b) Primer für die Sequenzierungsreaktion*

Sequenzierungsreaktion an einem einzelnen NSKF-Molekül wird durch ein Primer-Molekül ermöglicht. Ein sequenzspezifischer Primer ist erfindungsgemäß notwendig, um die Sequenzierungsreaktion jeweils an einer bestimmten / spezifischen Zielsequenz innerhalb der Gesamtsequenz durchführen zu können. Der für die Analyse einer SNP-Stelle, bzw. einer Zielsequenz einzusetzende sequenzspezifische Primer stellt eine Population von Primer-Molekülen mit identischer Struktur dar. Für die Analyse mehrerer, unterschiedlicher Zielsequenzen sind mehrere unterschiedliche Primer-Populationen notwendig.

Durch die Verwendung sequenzspezifischer Primer werden nur die relevanten Sequenzabschnitte, die Zielsequenzen, analysiert. Im erfindungsgemäßen Verfahren wird die zu sequenzierende Länge der Sequenzen möglichst niedrig gehalten, damit die Geschwindigkeit der Analyse steigt.

Ein sequenzspezifischer Primer bindet an eine für ihn spezifische Primerbindungsstelle in der zu analysierenden Sequenz, PBS. Die Zusammensetzung und die Länge der Primer werden für jede potentielle SNP-Stelle, bzw. Zielsequenz, optimiert. Beispiele für Optimierungsschritte sind in Rychlik et al. NAR 1990 v.18 S.6409 dargestellt. Bei der Primerwahl bzw. bei der Wahl der PBS (Primerbindungsstelle) sind folgende Aspekte besonders zu berücksichtigen:

1) Die zu analysierende SNP-Stelle sollte entweder gleich



nach dem 3'-Ende des Primers oder innerhalb der nächsten 2 bis 50 NTs, vorzugsweise 2 bis 20 NTs liegen .

- 2) Die Positionierung (die Wahl der Sequenzlänge und der Zusammensetzung) der PBS zu SNP-Stelle sollte so erfolgen, dass die verschiedenen PBS-Sequenzen und die korrespondierenden Primer-Sequenzen möglichst ähnliche ~~Annealing-Temperaturen~~ besitzen, um bei möglichst einheitlichen Hybridisierungsbedingungen zu binden. Das kann beispielsweise durch Veränderung der PBS-Position im Bezug auf die jeweilige, zu analysierende SNP-Stelle oder durch die Veränderung der Primersequenzlänge erfolgen (Rychlik et al. NAR 1990 v.18 S.6409).
- 3) Der minimale Abstand zwischen Primern, die an denselben Strang in der Gesamtsequenz binden, sollte die durchschnittliche NSKF-Länge nicht unterschreiten.

Es können Primer für beide Stränge einer Doppelstrang-NSK verwendet werden. Damit lassen sich beispielsweise nah aneinander liegende SNP-Stellen erfassen, oder man kann eine Kontrolle einer SNP-Stelle in beiden Strängen vornehmen.

Vorzugsweise beträgt die Länge des Primers zwischen 6 und 100 NTs, optimalerweise zwischen 10-30 oder 30-40 oder 40-50. Für verschiedene SNP-Stellen, bzw. Zielsequenzen können Primer mit unterschiedlicher Länge eingesetzt werden.

Für die SNP-Analyse mit sequenzspezifischen Primern werden Primer erfindungsgemäß in einer Hybridisierungslösung an die auf der Reaktionsoberfläche immobilisierten NSKFs hybridisiert (Annealing-Reaktion).

#### c) Immobilisierung von NSKFs

In dieser Ausführungsform werden erfindungsgemäß die NSKF-Primer-Komplexe ausschließlich über die NSKFs an die Oberfläche gebunden (direkte Bindung von NSKFs an die

Oberfläche), wobei die bereitgestellten NSKF-Moleküle an die plane Oberfläche in zufälliger Anordnung gebunden werden.

Die Immobilisierung der NSKFs erfolgt vorzugsweise an einem der beiden Ketten-Enden (s.o.). Die Immobilisierung kann auch durch eine unspezifische Bindung, wie z.B. durch Austrocknung der NSKFs enthaltenden Probe auf der planen Oberfläche erreicht werden. Die Dichte der Immobilisation kann zwischen 10 und 100, 100 und 10.000, 10.000 und 1.000.000 NSKFs pro  $100 \mu\text{m}^2$  liegen.

d) *Hybridisierung*

Die gebundenen NSKFs und die Primer werden unter stringenten Hybridisierungsbedingungen inkubiert, die eine möglichst selektive Anbindung (Annealing) der Primer an die entsprechenden Primerbindungsstellen der NSKFs erlauben. Optimale Hybridisierungsbedingungen hängen von der genauen Struktur der Primerbindungsstellen und der jeweiligen Primerstrukturen ab und lassen sich beispielsweise nach Rychlik et al. NAR 1990 v.18 S.6409 berechnen.

Die Primer stellen vorzugsweise ein Primergemisch dar. Die Konzentrationen einzelner sequenzspezifischer Primer (Einzelkonzentrationen von Primer-Populationen) liegen beispielsweise zwischen 10pmol/l und 1mmol/l, vorzugsweise zwischen  $0.1 \mu\text{mol/l}$  und  $10 \mu\text{mol/l}$ . Die Gesamtkonzentration von Primern im Primergemisch liegt vorzugsweise zwischen 1nmol/l und 10mmol/l. Das Verhältnis zwischen einzelnen Primer-Populationen kann variieren. Primer können in deutlichem Überschuss über die immobilisierten NSKFs zugegeben werden, so dass die Hybridisierungszeit gering ist.

Die für die Detektion notwendige Dichte von extensionsfähigen NSKF-Primer-Komplexen beträgt ca. 10 bis 100 pro  $100 \mu\text{m}^2$ . Sie kann vor, während oder nach der Hybridisierung der Primer erreicht werden.

Bei einer bekannten NSKF-Konzentration können in einer Ausführungsform die Immobilisierungsbedingungen so gewählt werden, dass die NSKFs in einer Dichte von ca. 10 bis 1000 Moleküle pro  $100\mu\text{m}^2$  gebunden werden. NSKFs bestimmen somit die Dichte der NSKF-Primer-Komplexe.

In einer anderen Ausführungsform kann die Dichte der immobilisierten NSKFs wesentlich höher als 1000 NSKFs pro  $100\mu\text{m}^2$  liegen, z.B. 1.000.000 pro  $100\mu\text{m}^2$ . Die für die optische Detektion notwendige Dichte der NSKF-Primer-Komplexe wird während der Primer-Hybridisierung erreicht. Dabei sind die Hybridisierungsbedingungen (z.B. Temperatur, Zeit, Puffer) so zu wählen, dass die Primer nur an einen Teil der immobilisierten NSKFs binden.

Bei unbekannter NSKF-Konzentration und entsprechend unbekannter Immobilisationsdichte kann die Hybridisierung (Annealing) von Primern an die NSKFs zu einer höheren als optimale Dichte von NSKF-Primer-Komplexen führen.

Aus diesem Grund wird in einer vorteilhaften Ausführungsform ein Teil der NSKFs enthaltenden Probe für die Ermittlung der optimalen Dichte verwendet. Dieser Teil wird auf einer Reaktionsoberfläche immobilisiert, die Primer werden an die NSKFs hybridisiert und die entstandenen NSKF-Primer-Komplexe werden durch den Einbau von Fluoreszenzfarbstoff tragenden NT\*s (z.B. Cy3-dCTP, Amersham Pharmacia Biotech) markiert. Aus der ermittelten Dichte lässt sich einerseits die eventuell notwendige Verdünnung oder Konzentrierung der ursprünglichen Probe für den endgültigen Sequenzierungsansatz errechnen (Die Hybridisierungsbedingungen werden beibehalten). Andererseits können daraus notwendige Veränderungen in den Hybridisierungsbedingungen errechnet werden, beispielsweise eine Verkürzung der Hybridisierungszeit, wobei die NSKF-Immobilisierungsdichte konstant bleibt.

Das Mengen-Verhältnis zwischen Primerpopulationen kann unterschiedlich oder gleich groß sein. Durch eine höhere Primerkonzentrationen können gewisse, beispielsweise seltenere Sequenzen mit größerer Wahrscheinlichkeit in einem bestimmten Zeitraum gebunden werden.

Der große Vorteil der beschriebenen Verfahrensanordnung gegenüber einer Verfahrensanordnung mit auf einer Oberfläche immobilisierten sequenzspezifischen Primern und einer anschließenden Hybridisierung von Proben an diese Primer ist die deutliche Verkürzung der Zeit für die Hybridisierung (Annealing) zwischen den sequenzspezifischen Primern und den zu analysierenden Proben auf der Reaktionsoberfläche.

#### Legenden zu Figuren 1 bis 8

##### Legende zu Fig. 1

##### Schematische Darstellung der Sequenzierung einer langen Nukleinsäurekette

Der Sequenzierung und der Rekonstruktion von langen Nukleinsäuresequenzen (NSKs) liegt das Shotgun-Prinzip zugrunde. Die Sequenz eines langen DNA-Stücks wird dabei durch die Sequenzierung kleiner Fragmente (NSKFs) und eine nachfolgende Rekonstruktion ermittelt.

- 1) Ausgangsmaterial - die zu analysierende lange Nukleinsäuresequenz, Gesamtsequenz
- 2) Fragmente von 50-1000 bp - die im Fragmentierungsschritt aus der Gesamtsequenz erzeugten NSKFs
- 3) Fragmente mit jeweils einem Primer - NSKF-Primer-Komplexe
- 4) Immobilisierte Fragmente - an die plane Oberfläche

gebundene NSKF-Primer-Komplexe, in dieser Ausführungsform erfolgt die Bindung am 3'-Ende der NSKFs

5) Zugabe einer Lösung mit Polymerasen und NT\*s - der erste Schritt in einem Zyklus der Sequenzierungsreaktion

6) Waschschrift - nach dem Einbauschritt wird die Oberfläche gewaschen

7) Detektion - die Signale von einzelnen eingebauten NT\*s werden detektiert

8) Entfernung der Markierung und der zur Termination führenden Gruppe

Legende zur Fig. 2

#### Beispiele für allgemeine Struktur von NSKF-Primer-Komplexen

In dieser Ausführungsform wird eine einheitliche Primerbindungsstelle (PBS) an das 3'-Ende der NSKFs angekoppelt und an diese PBS bindet ein einheitlicher Primer.

Legende zur Fig. 3

Ein Beispiel für die Ankopplung einer einheitlichen Primerbindungsstelle (PBS), die eine funktionelle Gruppe zur Bindung an die Oberfläche trägt.

In diesem Fall wird ein doppelsträngiger Oligonukleotidkomplex (3a), der beispielsweise eine Modifikation an beiden Strängen hat (3b), an die doppelsträngigen NSKFs liegt (3c). Nach Denaturierung entstehen einzelsträngige NSKFs mit einheitlicher PBS (3d).

Legende zur Fig. 4

**Ein anderes Beispiel für die Erzeugung einer einheitlichen Primerbindungsstelle (PBS).**

In diesem Fall werden NTs an das 3'-Ende der einzelsträngigen NSKFs angekoppelt (ein so genanntes „Tailing“). Durch Verwendung eines einheitlichen NT entsteht eine einheitliche PBS.

Legende zur Fig. 5

**Beispiel für die Bindung von NSKFs an eine gelartige Reaktionsoberfläche.**

Auf einer festen Unterlage (1) haftet eine Gelschicht (2), z.B. ein Polyacrylamidgel (Fig. 5a), oder haften viele Gelkügelchen (5), z.B. Agarosekügelchen (Fig. 5b). An die Oberfläche des Gels sind NSKFs (4) gebunden. Die NSKFs tragen eine funktionelle Gruppe, z.B. Biotin, und sind an das Gel über Streptavidin oder Avidin (3) gebunden.

Legende zur Fig. 6

**Beispiel für eine Durchflussvorrichtung**

Eine gelartige Reaktionsoberfläche (1) ist auf einer für das Anregungs- und Fluoreszenzlicht durchlässigen festen Unterlage (2) befestigt. Sie bilden zusammen den Deckel der Flow-Cell. Die Flüssigkeiten in der Flow-Cell können kontrolliert ausgetauscht werden, wobei die Flow-Cell zusammen mit Vorratsbehälter (3), Pumpe (4) und Ventil (5) eine Durchflussvorrichtung bilden. Auf der Reaktionsoberfläche sind NSKF-Primer-Komplexe gebunden (hier nicht abgebildet).

Die Signale der eingebauten NT\*s werden mit der Detektionsapparatur (6) detektiert.

Legende zur Fig. 7

**Schematische Darstellung der Analyse von mRNA-Population**

Der Analyse liegt die Sequenzierung kurzer Abschnitte von mRNA zugrunde.

- 1) mRNA - die zu analysierende mRNA-Population, in diesem Beispiel bestehend aus zwei unterschiedlichen mRNA-Molekülpopulationen (dünne und dicke Streifen repräsentieren mRNA-Moleküle)
- 2) Immobilisierte mRNA - an die plane Oberfläche gebundene mRNA-Primer-Komplexe, in diesem Beispiel erfolgt die Bindung durch die Oligo-dT-Primer
- 3) Zugabe einer Lösung mit Polymerasen und NT\*s - der erste Schritt in einem Zyklus der Sequenzierungsreaktion
- 4) Waschschrift - nach dem Einbauschrift wird die Oberfläche gewaschen
- 5) Detektion - die Signale von einzelnen eingebauten NT\*s werden detektiert
- 6) Entfernung der Markierung und der zur Termination führenden Gruppe

Legende zur Fig. 8

**Beispiel für ein Detektionssystem**

Dargestellt ist ein Weitfeld-Optik-Detektionssystem. Nach dem Einbau von markierten NT\*s wird die Oberfläche (7)

abgescannt, wobei die Fluoreszenzsignale von einzelnen, an die NTs gekoppelten Farbstoffmolekülen detektiert werden.

**Fig. 8a** Schematische Darstellung eines Abschnittes der Reaktionsoberfläche (grau), der abgescannt wird. Die Kreise entsprechen jeweils der Aufnahme eines 2D-Bildes und repräsentieren die Flächen, von denen die Fluoreszenzsignale detektiert werden. Dabei werden pro Aufnahme mehrere Signale (beispielsweise 100 bis 10.000) von einzelnen Molekülen gleichzeitig registriert. Die Reaktionsoberfläche wird in jedem Zyklus abgescannt, wobei während des Scannvorganges mehrere Aufnahmen von unterschiedlichen Stellen der Oberfläche gemacht werden. Dabei können bis zu mehreren Millionen Signale von eingebauten NT\*s aufgenommen werden. Die hohe Parallelität ist die Grundlage für die Geschwindigkeit des Verfahrens.

**Fig. 8b** Eine Aufnahme (ein 2D-Bild) mit Signalen von einzelnen, eingebauten NT\*s.

**Fig. 8c** Ausschnitt aus Abbildung 8b. Der Ausschnitt zeigt Signale von vier eingebauten NT\*s. Jedes Signal besitzt charakteristische Eigenschaften der Einzelmolekülsignale (s. Beschreibung) und kann auf grund dieser identifiziert werden (vorzugsweise mit Hilfe eines Computer-Programms). Jedem der identifizierten Signale werden die entsprechenden X,Y-Koordinaten zugeordnet.



Patentansprüche:

1. Verfahren zur parallelen Sequenzanalyse von Nukleinsäuresequenzen (Nukleinsäureketten, NSKs), bei dem man

Fragmente (NSKFs) einzelsträngiger NSKs mit einer Länge von etwa 50 bis 1000 Nukleotiden erzeugt, die überlappende Teilsequenzen einer Gesamtsequenz darstellen können, man

die NSKFs unter Verwendung eines einheitlichen oder mehrerer unterschiedlichen Primer in Form von NSKF-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der NSKFs unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

- a) zu den auf der Oberfläche gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs\*) enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs\* jeweils an den NTs\* befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs\* durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs\* an der 3'-Position strukturell so modifiziert sind, daß die Polymerase nach Einbau eines solchen NT\* in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT\* in denselben Strang einzubauen, wobei der zur Termination führende Substituent und der Fluoreszenzfarbstoff abspaltbar ist, man

- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT<sup>\*</sup> verlängert werden, man
- c) die in Stufe b) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs<sup>\*</sup> geeignet sind, man
- d) die einzelnen, in komplementäre Stränge eingebauten NTs<sup>\*</sup> durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- e) zur Erzeugung unmarkierter (NTs oder) NSKFs die zur Termination führenden Substituenten und die Fluoreszenzfarbstoffe von den am komplementären Strang angefügten NTs<sup>\*</sup> abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus nur jeweils ein markiertes NT\* einsetzt.
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils zwei unterschiedlich markierte NTs\* einsetzt.
4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils vier unterschiedlich markierte NTs\* einsetzt.
5. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die NSKs Varianten einer bekannten Referenzsequenz sind und man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei unterschiedlich markierte NTs\* und zwei unmarkierte NTs einsetzt und man die Gesamtsequenzen durch Vergleich mit der Referenzsequenz ermittelt.
6. Verfahren nach den Ansprüchen 1 bis 5, dadurch gekennzeichnet, dass man in die NSKFs jeweils eine Primerbindungsstelle (PBS) einführt, wobei man bei doppelsträngigen NSKs an beiden komplementären Einzelsträngen jeweils eine PBS einführt und wobei die Primerbindungsstellen für alle NSKFs jeweils gleiche oder verschiedene Sequenzen aufweisen.
7. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die NSKFs mit Primern in einer Lösung unter Bedingungen in Kontakt bringt, die zur Hybridisie-

rung der Primer an die Primerbindungsstellen (PBSs) der NSKFs geeignet sind, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen, und man die gebildeten NSKF-Primer-Komplexe anschließend auf der Reaktionsoberfläche bindet.

8. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die NSKFs zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit Primern unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der NSKFs geeignet sind, wobei NSKF-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.
9. Verfahren nach den Ansprüchen 1 bis 6, dadurch gekennzeichnet, dass man die Primer zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit NSKFs unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der NSKFs geeignet sind, wodurch NSKFs an die Oberfläche gebunden und NSKF-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.
10. Verfahren nach den Ansprüchen 1 bis 9, dadurch gekennzeichnet, dass die Dichte der extensionsfähigen NSKF-Primer-Komplexe zwischen etwa 10 und 100 pro  $100\mu\text{m}^2$  liegt.
11. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Nukleinsäuresequenzen (NSKs) Sequenzabschnitte einer Gesamtsequenz sind und die Primer sequenzspezifische Primer sind, wobei man

einzelsträngige NSKFs mit einer Länge von etwa 30 bis 1000 Nukleotiden bereitstellt, die überlappenden Teilsequenzen der Gesamtsequenz entsprechen, man

die NSKF-Moleküle direkt an einer planen Oberfläche in einer zufälligen Anordnung bindet, man

mit einer oder mehreren sequenzspezifischen Primerpopulationen eine Hybridisierung (Annealing) an die immobilisierten NSKFs durchführt, wobei die Dichte der einzelnen extensionsfähigen NSKF-Primer-Komplexe zwischen 10 und 100 pro  $100\mu\text{m}^2$  liegt, man

eine zyklische Aufbaureaktion der zu NSKFs komplementären Stränge durchführt, indem man

- a) zu den gebundenen NSKF-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, daß sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs' an der 3'-Position strukturell so modifiziert sind, daß die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der zur Termination führende Substituent und der Fluoreszenzfarbstoff abspaltbar ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man

- c) die in Stufe b) erhaltenen stationären Phase unter Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs' geeignet sind, man
- e) die einzelnen, in komplementäre Stränge eingebauten NT\*-Moleküle durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- e) die zur Termination führenden Substituenten und die Fluoreszenzfarbstoffe von den am komplementären Strang angefügten NTs' zur Erzeugung unmarkierter (NTs oder) NSKFs abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der zu Termination führenden Gruppen mit den Fluoreszenzfarbstoffen geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner NSKF-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser NSKFs durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt.

## 12. Verfahren zur hoch parallelen Analyse der Genexpression bei dem man

einzelsträngige Genprodukte bereitstellt, man

die Genprodukte unter Verwendung eines einheitlichen oder mehrerer unterschiedlichen Primer in Form von Genprodukt-Primer-Komplexen auf einer Reaktionsoberfläche in einer zufälligen Anordnung bindet, man

eine zyklische Aufbaureaktion des komplementären Stranges der Genprodukte unter Verwendung einer oder mehrerer Polymerasen durchführt, indem man

- a) zu den auf der Oberfläche gebundenen Genprodukt-Primer-Komplexen eine Lösung zugibt, die eine oder mehrere Polymerasen und ein bis vier modifizierte Nukleotide (NTs') enthält, die mit Fluoreszenzfarbstoffen markiert sind, wobei die bei gleichzeitiger Verwendung von mindestens zwei NTs' jeweils an den NTs' befindlichen Fluoreszenzfarbstoffe so gewählt sind, dass sich die verwendeten NTs' durch Messung unterschiedlicher Fluoreszenzsignale voneinander unterscheiden lassen, wobei die NTs' an der 3'-Position strukturell so modifiziert sind, daß die Polymerase nach Einbau eines solchen NT' in einen wachsenden komplementären Strang nicht in der Lage ist, ein weiteres NT' in denselben Strang einzubauen, wobei der zur Termination führende Substituent und der Fluoreszenzfarbstoff abspaltbar ist, man
- b) die in Stufe a) erhaltene stationäre Phase unter Bedingungen inkubiert, die zur Verlängerung der komplementären Stränge geeignet sind, wobei die komplementären Stränge jeweils um ein NT' verlängert werden, man
- c) die in Stufe b) erhaltene stationäre Phase unter

Bedingungen wäscht, die zur Entfernung nicht in einen komplementären Strang eingebauter NTs\* geeignet sind, man

- d) die einzelnen, in komplementäre Stränge eingebauten NTs\* durch Messen des für den jeweiligen Fluoreszenzfarbstoff charakteristischen Signals detektiert, wobei man gleichzeitig die relative Position der einzelnen Fluoreszenzsignale auf der Reaktionsoberfläche bestimmt, man
- e) zur Erzeugung unmarkierter, (NTs oder) Genprodukte die zur Termination führenden Substituenten und die Fluoreszenzfarbstoffe von den am komplementären Strang angefügten NTs\* abspaltet, man
- f) die in Stufe e) erhaltene stationäre Phase unter Bedingungen wäscht, die zur Entfernung der Fluoreszenzfarbstoffe und der Liganden geeignet sind, man

die Stufen a) bis f) gegebenenfalls mehrfach wiederholt,

wobei man die relative Position einzelner Genprodukt-Primer-Komplexe auf der Reaktionsoberfläche und die Sequenz dieser Genprodukte durch spezifische Zuordnung der in Stufe d) in aufeinanderfolgenden Zyklen an den jeweiligen Positionen detektierten Fluoreszenzsignale zu den NTs bestimmt und man aus den ermittelten Teilsequenzen die Identität der Genprodukte bestimmt.

13. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus nur jeweils ein markiertes NT\* einsetzt.

14. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass



- man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils zwei unterschiedlich markierte NTs\* einsetzt.
15. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in jedem Zyklus jeweils vier unterschiedlich markierte NTs\* einsetzt.
  16. Verfahren nach Anspruch 12, dadurch gekennzeichnet, dass bereits bekannte Gene als Referenzsequenzen dienen und man die Stufen a) bis f) der zyklischen Aufbaureaktion mehrfach wiederholt, wobei man in den Zyklen abwechselnd jeweils zwei unterschiedlich markierte NTs\* und zwei unmarkierte NTs einsetzt und man die Identität der Genprodukte durch Vergleich der gewonnenen Sequenzen mit denen der Referenzsequenzen ermittelt.
  17. Verfahren nach den Ansprüchen 12 bis 16, dadurch gekennzeichnet, dass man in die Genprodukte jeweils eine Primerbindungsstelle (PBS) einführt, wobei die Primerbindungsstellen für alle Genprodukte jeweils gleiche oder verschiedene Sequenzen aufweisen.
  18. Verfahren nach den Ansprüchen 12 bis 17, dadurch gekennzeichnet, dass man die Genprodukte mit Primern in einer Lösung unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der Genprodukte geeignet sind, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen, und man die gebildeten Genprodukt-Primer-Komplexe anschließend auf der Reaktionsoberfläche bindet.
  19. Verfahren nach den Ansprüchen 12 bis 17, dadurch gekennzeichnet, dass man die Genprodukte zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit Primern unter Bedingungen in Kontakt bringt, die zur

Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der Genprodukte geeignet sind, wobei Genprodukt-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.

20. Verfahren nach den Ansprüchen 12 bis 17, dadurch gekennzeichnet, dass man die Primer zunächst auf der Reaktionsoberfläche immobilisiert und erst anschließend mit Genprodukten unter Bedingungen in Kontakt bringt, die zur Hybridisierung der Primer an die Primerbindungsstellen (PBSs) der Genprodukte geeignet sind, wodurch Genprodukte an die Oberfläche gebunden und Genprodukt-Primer-Komplexe gebildet werden, wobei die Primer untereinander gleiche oder verschiedene Sequenzen aufweisen.
21. Verfahren nach den Ansprüchen 12 bis 20, dadurch gekennzeichnet, dass die Dichte der extensionsfähigen Genprodukt-Primer-Komplexe zwischen etwa 10 und 100 pro  $100\mu\text{m}^2$  liegt.
22. Verfahren nach Ansprüchen 1 bis 21 dadurch gekennzeichnet, dass der Fluoreszenzfarbstoff zusammen mit dem zur Termination führenden Substituenten abgespalten wird.
23. Verfahren nach Ansprüchen 1 bis 21 dadurch gekennzeichnet, dass zunächst der Fluoreszenzfarbstoff abgespalten wird und erst danach der zur Termination führende Substituent abgespalten wird.
24. Verfahren nach Ansprüchen 1 bis 21, dadurch gekennzeichnet, dass man im Detektionsschritt (d) folgende Detektionsarten einsetzt: Weitfeld-Epifluoreszenzmikroskopie, Laser-Scanning-

Fluoreszenzmikroskopie, TIRF-Mikroskopie.

25. Verfahren nach den Ansprüchen 1 oder 11, dadurch gekennzeichnet, dass es ein Verfahren zur SNP-Analyse ist und man einen sequenzspezifischen Primer zur Identifizierung jeder SNP-Stelle in der Gesamtsequenz verwendet.
26. Verfahren nach Anspruch 25, dadurch gekennzeichnet, dass die Zahl der parallel zu analysierenden SNP-Stellen größer als 2 ist und man für jede SNP-Stelle einen sequenzspezifischen Primer verwendet.
27. Verfahren nach den Ansprüchen 1 bis 26, dadurch gekennzeichnet, dass die Reaktionsoberfläche aus der Gruppe bestehend aus Silicon, Glas, Keramik, Kunststoffen, Gelen ausgewählt ist.
28. Verfahren nach Anspruch 27, dadurch gekennzeichnet, dass die Kunststoffe Polycarbonate oder Polystyrole oder Derivate derselben sind.
29. Verfahren nach Anspruch 27, dadurch gekennzeichnet, dass die Gele Agarose- oder Polyacrylamidgele oder Derivate derselben sind.
30. Verfahren nach Anspruch 29, dadurch gekennzeichnet, dass die Gele 1 bis 2 % Agarose-Gele oder 10 bis 15 % Polyacrylamid-Gele sind.
31. Verfahren nach den Ansprüchen 1 bis 30, dadurch gekennzeichnet, dass die Polymerase eine DNA-Polymerase ohne 3'-5'-Exonukleaseaktivität ist.
32. Verfahren nach Anspruch 31, dadurch gekennzeichnet, daß die Polymerase aus der Gruppe bestehend aus viralen DNA-Polymerasen vom Sequenase-Typ, bakteriellen

thermolabilen und thermostabilen DNA-Polymerasen, DNA-Polymerasen Beta aus Eukaryonten und Reversen-Transkriptasen ausgewählt ist.

33. Verfahren nach den Ansprüchen 1 bis 32, dadurch gekennzeichnet, dass die Fluoreszenzfarbstoffe aus der Gruppe bestehend aus Cyanin-Farbstoffen, Rhodamine, Xanthene, Porphyrine und deren Derivaten ausgewählt sind.
34. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 33, dadurch gekennzeichnet, dass auf seiner Oberfläche die Nukleinsäureketten oder deren Fragmente in einer zufälligen Anordnung immobilisiert sind, wobei die Dichte der immobilisierten Nukleinsäureketten-Moleküle oder deren Fragmente zwischen 10 und 100 pro  $100\mu\text{m}^2$  liegt.
35. Träger zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 33, dadurch gekennzeichnet, dass auf seiner Oberfläche die Nukleinsäureketten oder deren Fragmente in einer zufälligen Anordnung immobilisiert sind, wobei die Dichte der immobilisierten Nukleinsäureketten-Moleküle oder deren Fragmente zwischen 100 und 1.000.000 pro  $100\mu\text{m}^2$  liegt.
36. Kit zur Durchführung des Verfahrens nach den Ansprüchen 1 bis 35, dadurch gekennzeichnet, daß es eine Reaktionsoberfläche (einen festen Träger), zur Durchführung des Verfahrens erforderliche Reaktionslösungen, ein oder mehrere Polymerasen, und Nukleotide (NTs) enthält, von denen ein bis vier mit Fluoreszenzfarbstoffen markiert sind, wobei die markierten NTs ferner strukturell so modifiziert sind (NT\* bzw. NTs\*), daß die Polymerase nach Einbau eines solchen NT\* in einen wachsenden komplementären Strang nicht in der Lage

ist, ein weiteres NT\* in denselben Strang einzubauen, wobei der zur Termination führende Substituent mit dem Fluoreszenzfarbstoff abspaltbar ist

37. Kit nach Anspruch 36, dadurch gekennzeichnet, daß es ferner Bestandteile enthält:

- a) zur Erzeugung von Einzelsträngen aus Doppelsträngen erforderliche Reagenzien,
- b) Nukleinsäuremoleküle, die als PBS in die NSKFs eingeführt werden,
- c) Oligonukleotid-Primer,
- d) zur Abspaltung der Substituenten mit den Fluoreszenzfarbstoffen erforderliche Reagenzien,
- e) Waschlösungen.

38. Kit nach den Anspruch 36 oder 37, dadurch gekennzeichnet, daß die Reaktionsoberfläche aus der Gruppe bestehend aus Silicon, Glas, Keramik, Kunststoffen, Gelen ausgewählt ist.

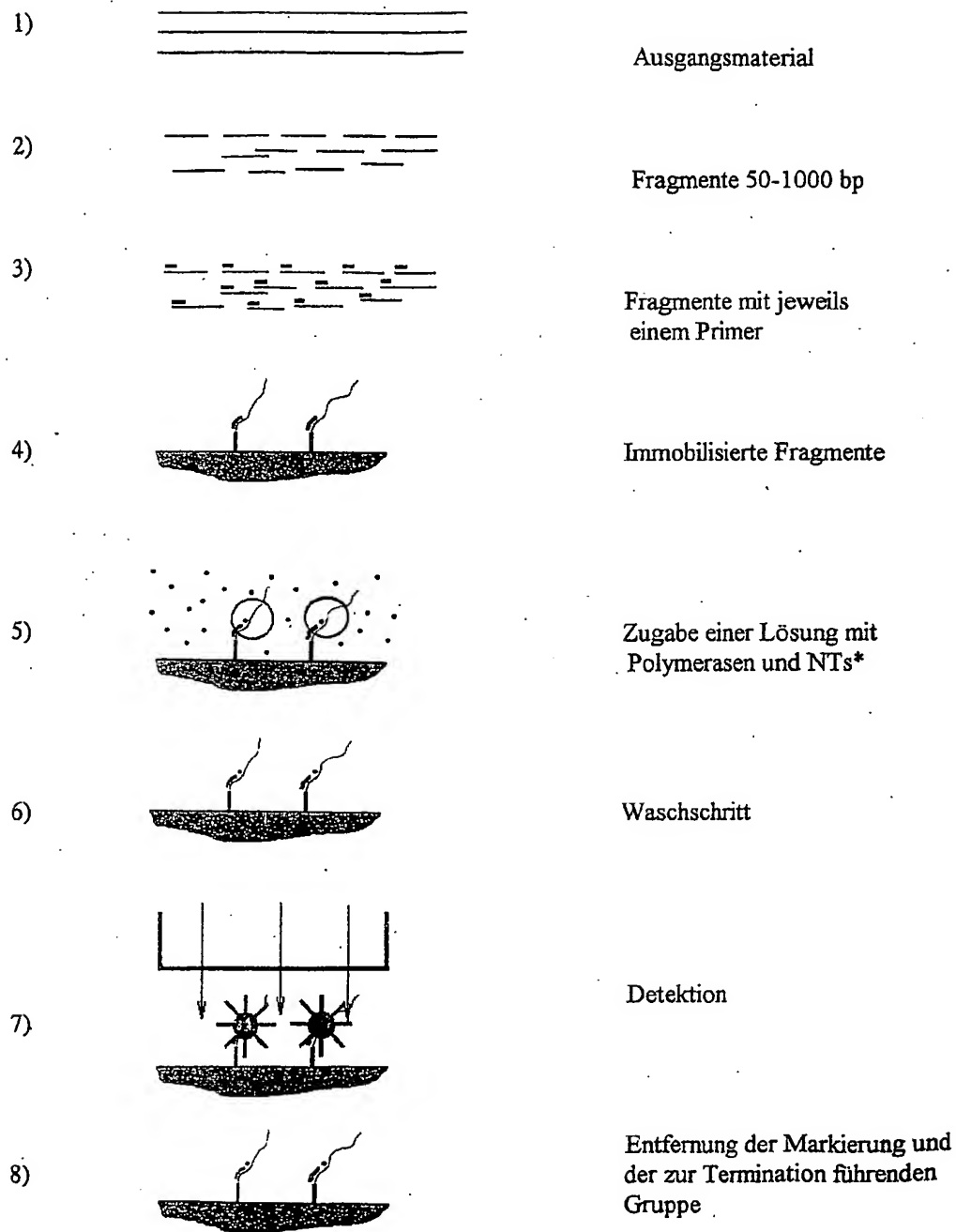
39. Kit nach Anspruch 38, dadurch gekennzeichnet, daß die Gele Polyacrylamidgele sind.

40. Kit nach Anspruch 39, dadurch gekennzeichnet, daß die Gele 5 bis 30% Polyacrylamid-Gele sind.

41. Kit nach den Ansprüchen 36 bis 40, dadurch gekennzeichnet, daß die DNA-Polymerase eine DNA-Polymerase ohne 3'-5'-Endonukleaseaktivität ist.

42. Kit nach den Ansprüchen 36 bis 41, dadurch gekennzeichnet, daß die an die Nukleotide gekoppelten Fluoreszenzfarbstoffe aus der Gruppe bestehend aus Cyanin-Farbstoffen, Rhodamine, Xanthene, Porphyrine und deren Derivaten ausgewählt sind.

Fig. 1

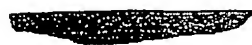
**Legende:**

feste Oberfläche

Polymerase

NT\*

Signal von NT\*

Mikroskop und  
Beleuchtung

**Fig. 2**

**Allgemeine Struktur für immobilisierte Nukleinsäurekettenfragmente**



**Legende:**

- 1) Primer
- 2) Primerbindungsstelle
- 3) einzelsträngiger Teil des Nukleinsäurekettenfragments, stammt vom Ausgangsmaterial

Fig. 3

### Oligonukleotid-Konstruktion mit einer Primerbindungsstelle und blockierten Enden.

3a)

Oligonukleotidkomplex      5'-----3'  
                                      3'//-----//5'

Primerbindungsstelle      //-----//

3b)

Modifikation einer Seite des Oligonukleotides an beiden Strängen. Nur ein Ende steht für die Ligation zur Verfügung.

xxxxx-5'-----3'  
       kkk-3'//-----//5'

Die modifizierenden Gruppen "xxxxx-" und "kkk-" können z.B. zur Immobilisation oder Markierung dienen.

3c)

Resultat der Ligation des Oligonukleotides mit den DNA-Fragmenten

xxxxx-5'-----//-----//3'-kkk  
       kkk-3//-----//-----5'-xxxxx

3d)

Konstruktion der DNA-Fragmente nach der Denaturierung

xxxxx-5'-----//-----//3'-kkk



Fig. 4

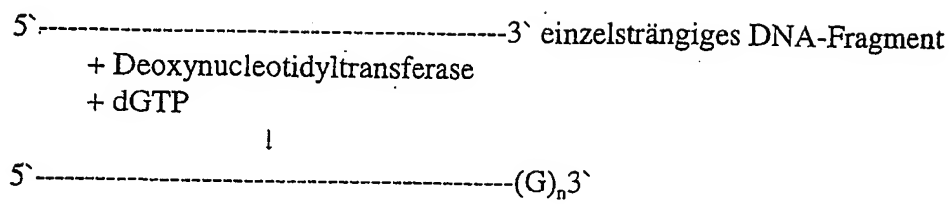
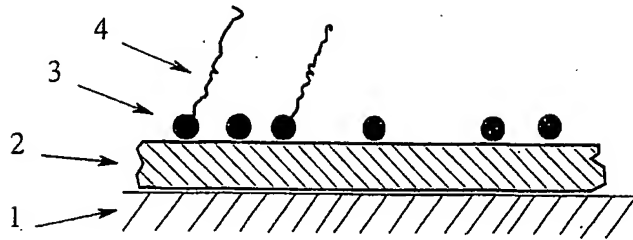
**Nukleotid-Tailing**

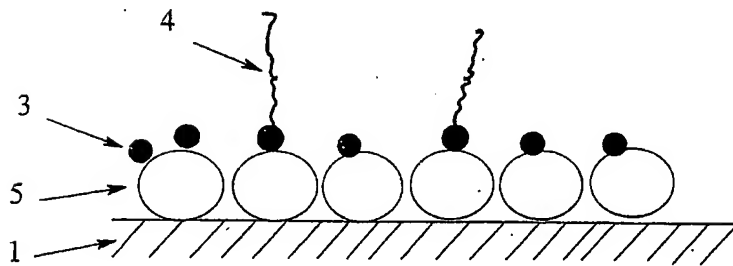


Fig. 5

5a)



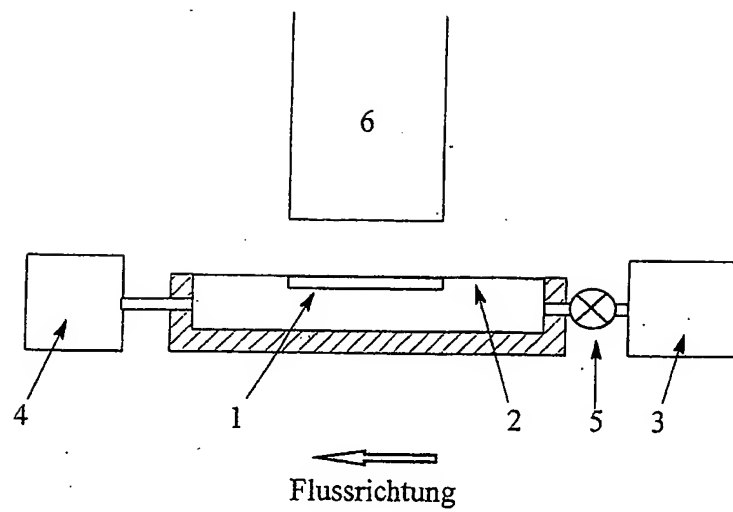
5b)



**Legende:**

- 1) feste Oberfläche
- 2) Gel
- 3) zur Immobilisation geeignete Gruppen, z.B. Avidin oder Streptavidin, oder Oligonukleotide
- 4) NSKFs
- 5) Agarosekügelchen

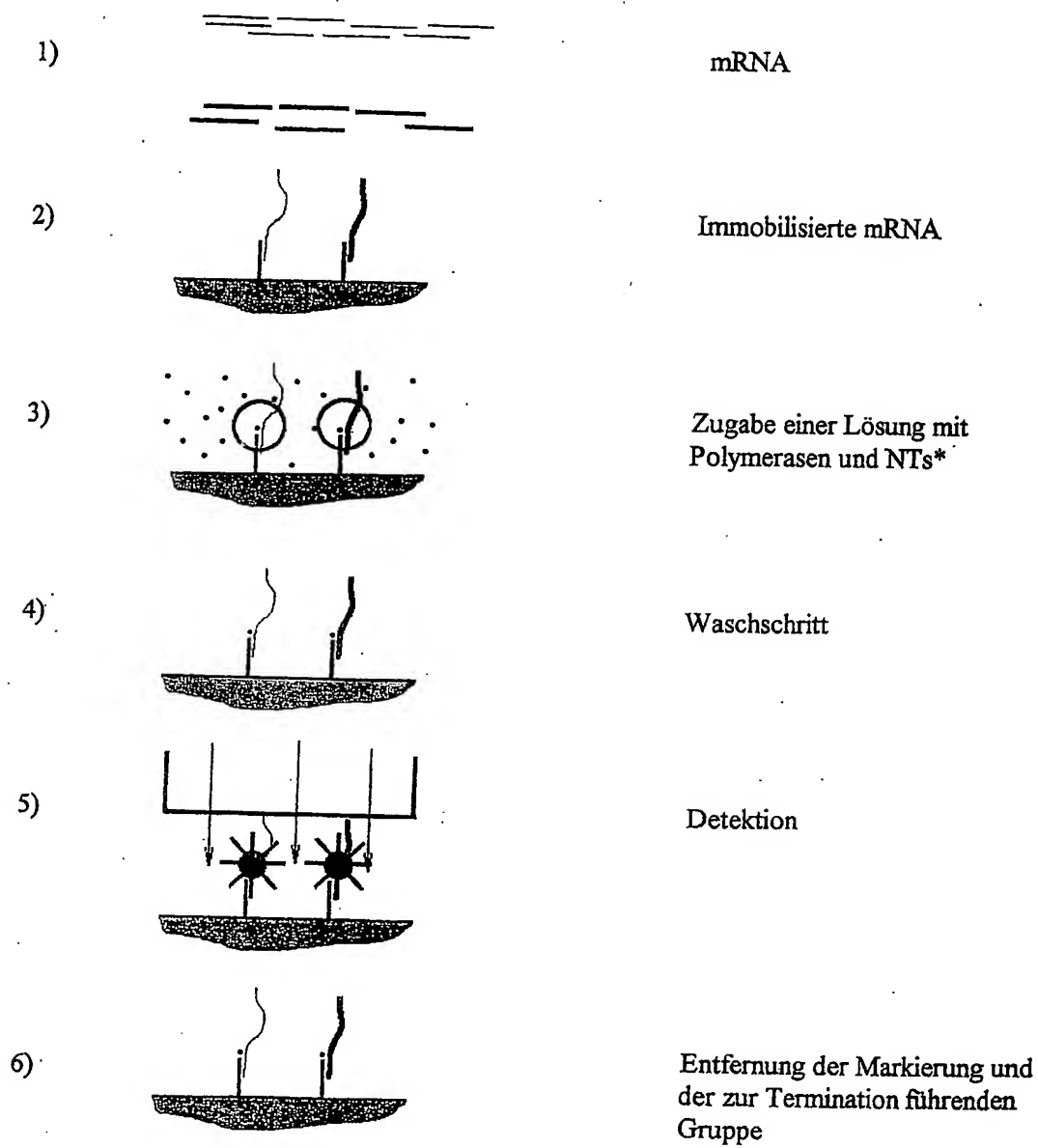
Fig. 6



**Legende:**

- 1) Reaktionsoberfläche
- 2) feste Unterlage
- 3) Vorratsbehälter
- 4) Pumpe
- 5) Ventil
- 6) Objektiv des Mikroskops

FIG. 7



**Legende:**

feste Oberfläche  
mit oligo-dT-Primern

Polymerase

NT\*

Signal von NT\*

Mikroskop und  
Beleuchtung

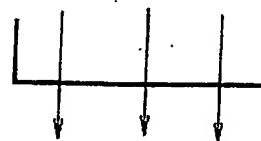
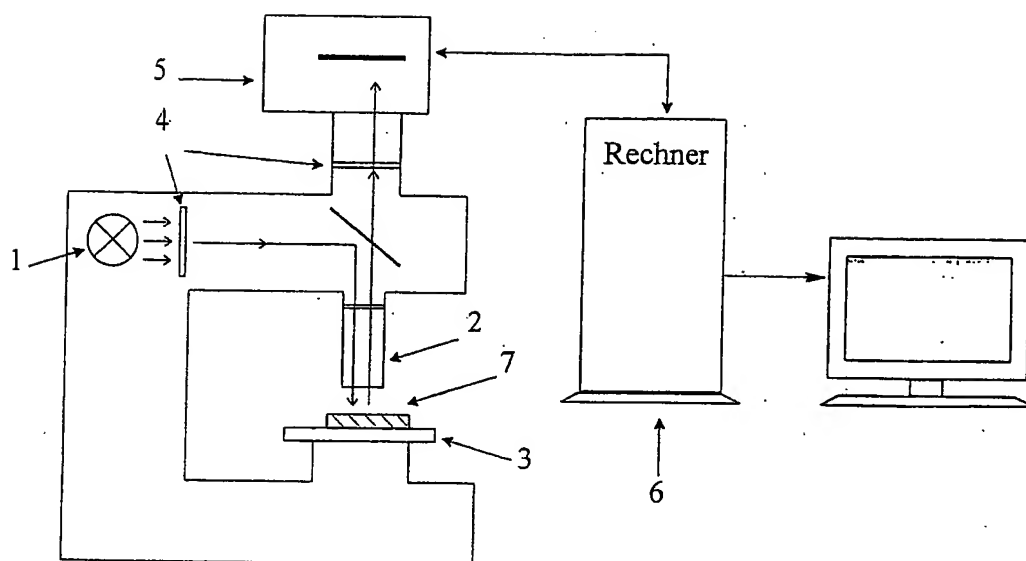


Fig. 8



**Legende:**

- 1) Lichtquelle zur Anregung der Fluoreszenz
- 2) Lichtleitender Teil
- 3) Scantisch
- 4) Vorrichtung zur Selektion von Spektren
- 5) Detektionsvorrichtung
- 6) Computer mit Steuerungs- und Analysefunktionen
- 7) Oberfläche mit immobilisierten NSKFs

